1997

# On the effect of the ethernet protocol in LAN Traffic modelling

I. Gusti Paramajaya
*University of Wollongong*

A Thesis entitled

# On the Effect of the Ethernet Protocol in LAN Traffic Modelling

Submitted to the
University of Wollongong
in partial fulfilment of the requirements for the degree of

**Master of Engineering**

in

**Telecommunications Engineering**

(Honours)

by

I Gusti Putu Bagus Paramajaya
Sarjana Teknik Elektro
Bandung Institute of Technology 1991

Department of Electrical and Computer Engineering
March 1997

# Abstract

There are two groups of traffic models that usually used in telecommunication network design, planning and performance evaluation. They are aggregate and individual models. In case of individual models, measurement-based studies are the most common approach used to obtain an accurate traffic model. The common assumption used in this approach is that the data link layer protocol (i.e., layer 2 of OSI Reference Model) does not significantly affect traffic characteristics of its higher layer where the traffic characteristics are of interest. In other words, this hypothesis assumes that traffic characteristics at the input of the data link protocol is statistically the same as the traffic characteristics at its output, i.e., the physical layer where the measurement can be easily done. This hypothesis is addressed as the null hypothesis.

This thesis is concerned with testing the validity and investigating the sensitivity of the null hypothesis against several parameters. For these purposes, an Ethernet network simulation model is developed and several statistical methods are utilised. Several simulation network configurations are considered as test configurations which represent different combination of test parameters, namely network parameters, protocol parameters and user parameters. Those configurations are used

to test the null hypothesis under two different traffic source models, i.e., Poisson and Pareto traffic models. Statistical test results indicated that the validity of the null hypothesis mostly depends on traffic model under consideration and is sensitive to the activity level of monitored station. The comprehensive statistical test results concluded that the null hypothesis is valid for relatively low-level activity stations (stations having packet departure rate up to 10 packets/second) and sensitive only to the user parameter of user activity level.

In addition, this thesis also presents a case study that demonstrates the liability of applying the null hypothesis in measurement-based traffic modelling without further formal validation.

# Acknowledgements

I would like to thank my supervisor Mr. David J. Atkinson for his helpful support, guidance and encouragement. I would also like to thank all staff in Department of Electrical and Computer Engineering University of Wollongong, whose supports made it possible for me to finish my study. In particular I would like to acknowledge the helpful statistical discussion and suggestion given by Dr. Pam J. Davy from Department of Applied Statistics through Statistical Consulting Services.

I am very grateful to my company PT TELEKOMUNIKASI INDONESIA for the financial support. I also thank my friends and colleagues Kalamullah Ramli, Chamilla Fernando, Narasimhan Sriram, Nagaraja, Li Xue, Ali Yazdian Varjani, Ery Yusuf, Widi Nugroho and Winahyo Pratomo.

My special thank goes to my wife who has faithfully stood beside me during problems and inconvenience during the thesis finalisation process in our first year of marriage. My father, mother, brother and sisters and relatives in my country also deserve thanks for their support and encouraging me to do my best.

Finally, I'd like to thank deeply all hidden supports given by my spiritual guidance Bhagawan Sri Sathya Sai Baba. Without Him, it is impossible for me to finish all my work.

IGPB Paramajaya

# Statement of Originality

The work described in this thesis is entirely my own, except where due reference is made in the text.

No work in this thesis has been submitted for a degree to any other university or institution

Signed

IGPB Paramajaya

# Table of Contents

# List of Abbreviations

| | |
|---|---|
| ATM | Asynchronous Transfer Mode |
| B-ISDN | Broadband Integrated Service Digital Network |
| CDF | Cumulative Distribution Function |
| CSMA/CD | Carrier-Sense Multiple Access with Collision Detection |
| CV | Coefficient of Variation |
| DES | Discrete Event Simulation |
| E30 | Configuration of 30 Stations with Exponential Packet IAT |
| E50 | Configuration of 50 Stations with Exponential Packet IAT |
| ECDF | Empirical Cumulative Distribution Function |
| EDF | Empirical Distribution Function |
| FCFS | First Come First Serve |
| FIFO | First In First Out |
| FRPP | Fractal Renewal Point Process |
| H | Hurst Parameter |
| Ho | The Null Hypothesis |
| IAT | Inter-Arrival Time |
| IDC | Index of Dispersion for Count |
| IDE | Integrated Development Environment |

| | |
|---|---|
| IDT | Inter-Departure Time |
| IP | Internet Protocol |
| ITA | Internet Traffic Archive |
| K-S | Kolmogorov-Smirnov |
| LAN | Local Area Network |
| LB | Large Buffer |
| LN | Long Network |
| LRD | Long-Range Dependence |
| LTCDF | Less Than Cumulative Distribution Function |
| MAC | Medium Access Control |
| MTCDF | More Than Cumulative Distribution Function |
| OSI | Open Systems Interconnection |
| P1 | Packet Size Distribution of Type 1 |
| P30 | Configuration of 30 Stations with Pareto Packet IAT |
| P50 | Configuration of 50 Stations with Pareto Packet IAT |
| Q-Q | Quantile-Quantile |
| SB | Small Buffer |
| SN | Short Network |
| SRD | Short-Range Dependence |
| TCP | Transmission Control Protocol |
| UI | Utilisation Index |
| VTP | Variance Time Plot |
| P2 | Packet Size Distribution of Type 2 |

# List of Figures

# List of Tables

# 1. Introduction

## 1.1 Background

Traffic modelling plays an important role in telecommunications network design, planning and performance evaluation. In analytical or simulation studies, the use of different traffic models may produce different results. Traffic modelling thus becomes an important issue in telecommunications network engineering.

Closely related to the subject of traffic modelling is traffic measurement that has been evolved for more than two decades. Traffic measurements are usually used to validate some types of traffic models (Pawlita 1989). Such models are in turn employed in two fundamental ways: analytical calculation or simulation study.

### 1.1.1 Link-Level (Aggregate) Model and Protocol-Specific (Individual) Model

Traffic models can be classified into two groups: link-level models and protocol-specific models (Paxson & Floyd 1994). The first group is an aggregate or cross-traffic model that is mainly useful to study the performance of interconnected networks (Cinotti et al. 1994; Cinotti et al. 1995; Fowler & Leland 1991). The models suggested in (Leland et al. 1993; Cinotti et al. 1994; Willinger et al. 1995) are counted to this group. The second group is the individual models. This includes models suggested in (Danzig et al. 1992; Jain & Routhier 1986; Liu, Anido & Chicharo 1994). This group of model is particularly important in evaluating the performance of local area network (LAN), such as ATM LAN (Chao et al. 1994; Newman 1994b) or in studying congestion management (Liu, Anido & Chicharo 1994). In many simulation studies, such as investigating the performance of particular protocol or gateway scheduling algorithm on network traffic, this protocol-specific modelling is of great important (Paxson & Floyd 1994).

### 1.1.2 Traffic Model and OSI Model

Figure 1.1 depicts the relationship between IEEE 802 reference model and OSI model (Stallings 1994, p. 407). It is assumed here that the reader is familiar with the concept and terminology of local area network (LAN) and internetworking (Comer 1991 pp. 51-59; Stallings 1994, pp. 471-538; Tanenbaum 1988, pp. 320-349). In an interconnected LAN, two hosts from two different LANs may communicate with each other through a device called a bridge. These two hosts may have the same network protocol (e.g., IP) but working in different LANs (i.e., one is in an Ethernet LAN and the other is in a Token Bus LAN). In traffic modelling, this fact has facilitated the assumption that the individual traffic model may be well represented by the network layer traffic characteristics. The measurement-based traffic modelling studies thus collected the traffic data from physical layer (aggregate traffic data) and analysed the statistics to determine an appropriate model.

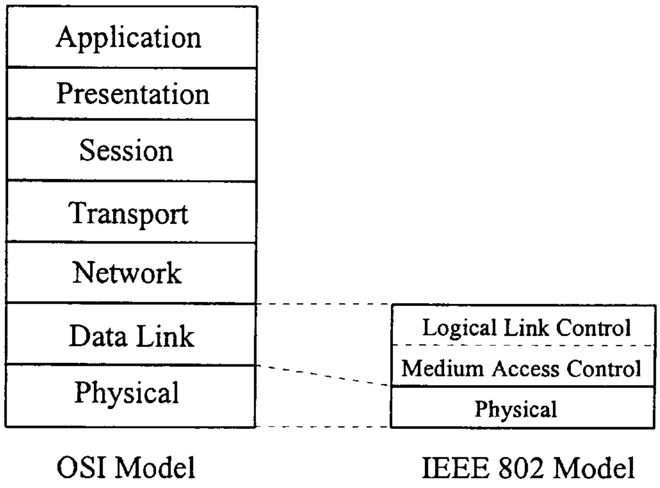


Figure 1.1 IEEE 802 Reference Model Relationship to OSI Model

In case of individual traffic models, data of an individual network user (mostly time stamp, packet size and source-destination address) is extracted from the aggregate traffic. The statistics from this individual traffic data are analysed and a model for individual network user (i.e., network protocol traffic characteristic) is proposed based on this analysis. In this case, the basic assumption that LAN protocol operation does not affect the traffic stream from higher layer has been implicitly used.

There are a number of problems associated with this approach of traffic modelling. The aggregate traffic data is the traffic stream resulting from the aggregation of many users. This traffic stream is affected by individual characteristics of the users as well as the network protocols used to manage the stream. The other problem is the measurement system that determines the accuracy of the collected traffic data (mostly the packet time stamp). Most of the traffic data does not produce true statistics due to the accuracy limitations of data collection system. Given that one may have very accurate traffic data from measurement, it is still necessary to investigate the validity and sensitivity of the basic assumption used in deriving individual source traffic model before that basic assumption can be justified. Thus this thesis is concerned with designing experimental study to investigate the validity and sensitivity of the basic assumption used in previous work on individual (source) traffic modelling. The numerical method of simulation is used as a tool for these purposes.

## 1.2 Thesis Outline

The thesis is composed of 7 chapters. They are organised as follows.

Chapter 2 reviews previous work in traffic modelling, especially in individual (source) traffic modelling. The basic assumption used in previous studies is identified as point of departure for the work in this thesis. This basic assumption is then addressed as the null hypothesis that needs to be tested. It is argued that there are 3 sets of parameters (i.e., network, protocol and user parameters) that may affect the null hypothesis and required further investigation. Following this identification, a problem in traffic measurement is presented. The problem is mainly dealt with time stamp accuracy. This problem may contribute as serious source of error in statistical calculation. The presence of this error strengthens the need of the work in this thesis. Chapter 2 is closed by arguing that numerical method of simulation is a suitable evaluation technique to conduct an experimental study for the purposes of this thesis. The simulation model and related statistical methods are described in the following chapters.

Chapter 3 mainly deals with the development, verification and validation of an Ethernet network simulation model. The Ethernet protocol has been considered as a test data link layer protocol mainly because most of the measurement-based studies considered in Chapter 2 were based on traffic data collected from Ethernet networks. The simulation model is designed to imitate the behaviour of real Ethernet network with some assumptions to simplify the model development. The model is designed based on a previously proposed CSMA/CD network simulation model (Sadiku & Ilyas 1995) with several improvements to account for the standard Ethernet protocol. The resulting simulation data will be used to test the validity and sensitivity of the null hypothesis by means of statistical methods described in Chapter 4.

Chapter 4 outlines several graphical and formal methods that are commonly used in testing the null hypothesis concerning with distribution of two sample data. The graphical methods of empirical cumulative distribution function (EDF for short) and Quantile-Quantile (Q-Q) plot are used as informal method in conjunction with formal methods. These informal methods will be useful to visually clarify the results of formal methods. All the results of testing the validity and sensitivity of the null hypothesis will be mainly based on formal numerical method known as Kolmogorov-Smirnov (K-S) test, which is commonly used to test continuous distribution data. We consider the K-S test with V statistic in testing the null hypothesis since it is more accurate to detect any departure in the tail of two distributions compared to the well-known D statistic. However, both statistics are calculated in every test and D statistic results serve as a reference.

The remainder of Chapter 4 describes second order statistical tests that are mainly used in studying the characteristic of aggregate traffic data in relation to the self-similarity issue. The self-similarity test results will be useful as a reference when comparing K-S test results of simulation data generated by different traffic models. All the statistical methods described above are utilised in the following two chapters.

The simulation model and the statistical methods described in previous chapters are employed extensively in Chapter 5. The experimental procedures used in the main part of this simulation study are firstly described in detail. Several simulation network configurations are considered as test configurations that represent different combination of test parameters as discussed in Chapter 2. Those configurations are used to test the null hypothesis under two different traffic source models, i.e., Poisson and Pareto traffic models. The statistical test results are presented either graphically or tabularly. Chapter 5 is closed by summarising the null hypothesis test results. This actually concludes the goal of this thesis, testing the validity of the null hypothesis and investigating its sensitivity against several parameters. However, to demonstrate the liability of the use of the null hypothesis in a previous traffic modelling study, we present a case study in Chapter 6.

Chapter 6 presents a case study of traffic measurement and modelling. Its aim is to demonstrate that the use of the null hypothesis in measurement-based traffic modelling without further formal validation may lead to the wrong conclusion for the derived traffic model.

Finally, Chapter 7 describes major results of the thesis and discusses the lessons learned from the whole investigation. It also presents some suggestions for possible improvements of this work as well as proposes scope for further research.

## 1.3 Contributions

This section lists the contributions contained in this thesis, along with the section where the relevant work is discussed.

1) Recognition of the unjustified basic assumption used in the previous study of traffic modelling that essentially assumed that the data link protocol does not affect the traffic stream from higher layer. This assumption has been addressed as the null hypothesis. See Section 2.3.

2) Recognition that the accuracy of Bellcore traffic data is poor even though it is claimed as a result of a high-time resolution measurement system. See Section 2.5.

3) Development of a standard Ethernet simulation model that is capable to imitate real Ethernet network behaviour in star network configuration and suitable for the purposes of testing the null hypothesis. See Chapter 3.

4) Designing experimental studies and its procedures to test the null hypothesis. This includes combining the knowledge of statistical methods and computer communication background. See Chapter 4 and Section 5.2.

5) Demonstrating by means of discrete event simulation that the aggregate traffic data generated by the Pareto traffic sources with shape parameter $\beta < 2$ (i.e., having infinite variance) exhibits long-range dependence or self-similar behaviour. See Section 5.4.2.

6) Demonstrating by means of discrete event simulation that the estimated Hurst parameter of Pareto aggregate traffic data become small as its shape parameter increases (i.e., as its variability reduces). See Section 5.4.2, Table 5.4.

7) Finding that under given conditions of simulation configurations (i.e., number of stations, bus length, station packet buffer and packet size distribution) the validity and the sensitivity of the null hypothesis depends on the traffic model under consideration. See Section 5.5 and 5.6.

# 2. LAN Traffic Source Modelling

## 2.1 Outline

This chapter reviews previous studies of LAN traffic modelling, especially measurement-based traffic modelling. This review identifies an unjustified basic assumption used in these studies which is later considered as the null hypothesis. The hypothesis that the numerical method of simulation is the suitable evaluation technique is also presented.

In Section 2.2, we identify methods used in previous studies as a point of departure for the work in this thesis. The basic assumption that data link protocol does not affect traffic stream characteristics from its higher layer is identified. This basic assumption is then addressed as the null hypothesis in Section 2.3. This hypothesis will be tested and investigated to find out its validity as well as its sensitivity against several parameters. Discussion on parameters that may affect the validity and sensitivity of the null hypothesis is presented in Section 2.4.

The importance of testing the null hypothesis is emphasised in Section 2.5 by demonstrating the problem in LAN traffic measurement. The rationale of using simulation technique as the suitable tool for the purposes of this thesis is presented in Section 2.6. Section 2.7 then concludes the chapter by briefly discussing related analytical studies and self-similarity issues in traffic modelling. This discussion is important to provide additional technical background for this study.

## 2.2 Measurement-Based LAN Traffic Modelling

Due to its existing need, LAN interconnection services are likely to be a dominant and an immediate application in future broadband networks. This has motivated a growing interest in the area of LAN traffic modelling. The aim is to find more accurate traffic models, since they are essential in performance evaluation and control of future broadband networks (Liu, Anido & Chicharo 1994).

The need of more accurate traffic models is also motivated by the discrepancy between characteristics of currently considered formal models for packet traffic (e.g., pure Poisson or Poisson-related models) and the observed characteristics of real LAN traffic (Leland et al. 1993). The presence of self-similar or fractal-like behaviour in real LAN traffic can not be explained by traditional traffic source models. This in turn has driven a vast study in source traffic modelling based on real traffic data. A number of studies (Cinotti et al. 1994; Gusella 1991; Jain & Routhier 1986; Leland et al. 1994) have suggested that the measurement-based traffic source models should be used to solve problems in network engineering since the ideal traffic source models could underestimate the important queuing statistics of the network.

There are several methods commonly used to derive traffic models from real traffic data. Most of these methods are based on idea of matching or fitting real traffic data to some known theoretical models using a variety of parameters (Addie, Zukerman & Neame 1995; Arlitt & Williamson 1996; Basu, Mukherjee & Klivansky 1996; Càceres et al. 1991; Cinotti et al. 1994; Leland et al. 1994; Liu, Anido & Chicharo 1994; Paxson & Floyd 1994; Robert & Lambrecht 1995). The idea of matching or fitting is reasonable and obvious since a good traffic model should have the following characteristics (Stamoulis, Anagnostou & Georgantas 1994): proximity to real sources, ease of fitting to real sources and appropriateness for modelling aggregate traffic.

In the case of traffic source models, the commonly used method is the fitting of real traffic data to some known distribution functions. This technique employs some statistical methods known as goodness-of-fit techniques (D'Agostino & Stephens 1986) that will be described more detail in Chapter 4. As an object and point of departure of this thesis, in particular we consider the previous study of LAN traffic characterisation and modelling (Liu, Anido & Chicharo 1994). This study will also be an object of our case study as presented in Chapter 6.

## 2.3 The Null Hypothesis

Liu, Anido and Chicharo (1994) studied Ethernet traffic data collected from a typical departmental LAN. The data is grouped according to its source-destination pair by extracting information from aggregate traffic data. The resulting inter-event distribution was plotted and fitted using the so-called Bond distribution. This distribution (Bond 1987) was previously used to model the error performance of a transmission link (Sexton & Reid 1992, pp. 163-169). This inter-event distribution was considered as an individual traffic source model (i.e., layer 3 traffic model). This study thus had implicitly assumed that the Ethernet protocol does not significantly affect the traffic characteristic of the network layer, e.g., TCP/IP. In other words, the null hypothesis **Ho**, that the distribution of traffic at physical layer (layer 1) is the same with the distribution of traffic from network layer (layer 3)(see Figure 1.1), has been used. Figure 2.1 illustrates this null hypothesis in a typical Ethernet network. The distribution of individual traffic $F_1$ extracted from aggregate traffic is considered the same with $F_2$, distribution of traffic from layer 3 or client of Ethernet protocol (e.g., TCP/IP); thus assuming that the Ethernet protocol (layer 2) has no significant effect on $F_2$. In this thesis, it is this assumption that will be tested and investigated against several parameters.

The same assumption has been used in other studies as well. Based on the traffic data collected from a token ring network, (Jain & Routhier 1986) proposed a traffic model called packet train model. Other protocol-specific traffic models have been proposed based on the same assumption (Danzig et al. 1992, Paxson & Floyd 1994).

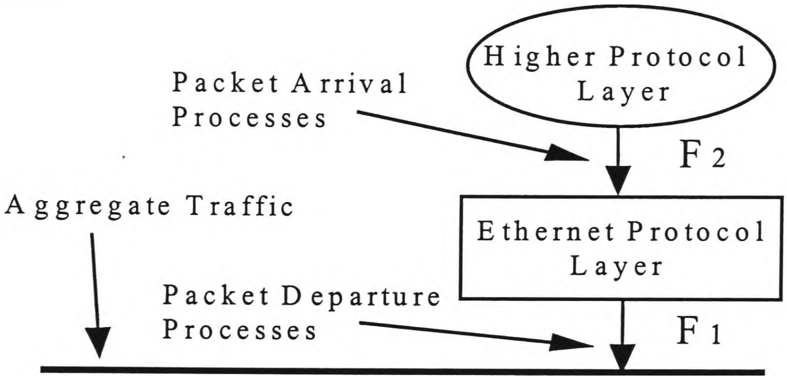


Figure 2.1 The Null Hypothesis in Ethernet Traffic Modelling (**Ho**: $F_1 = F_2$)

## 2.4 Factors Affecting the Null Hypothesis in an Ethernet Network

In Ethernet networks, each user has to contend to acquire the common channel. (More detail explanation of operation of Ethernet protocol will be described in Chapter 3). Thus, one may argue that statistics of aggregate traffic will obviously depend on several factors.

We identify three set of parameters that may affect the null hypothesis in Ethernet networks. These include network parameters, user parameters and protocol parameters. All of these factors may affect the packet transmission characteristic of an individual station. Since the most important statistic from real measured traffic is packet's time stamp, any factor that affects packet transmission characteristic will obviously affect this statistic.

As a random access technique (Schwartz 1987 pp. 441-447), The Ethernet protocol has some complex mechanisms to manage packet transmission processes. These mechanisms include carrier sensing, deference, collision detection and enforcement, backoff algorithm, buffering and jamming (Wolfberg 1983 pp. 10-122; Black 1993 pp. 125-130). Therefore, different characteristics may be observed in packet inter-departure time data irrespective of characteristics of its inter-arrival time.

### 2.4.1 Network Parameters

We consider the number of terminals (users) attached to the bus and the bus length as network parameters. The greater the number of active users, the more collisions may be expected and the packet transmission delay will increase. One packet may require several transmission attempts before it is transmitted successfully (without collision). Therefore, more collisions will influence packet transmission characteristic of each station and in turn will affect packet inter-departure time. In their study, Liu, Anido and Chicharo (1994) have used packet inter-departure time as the fundamental statistics to derive their traffic source model. In fact, earlier study (Smith & Kain 1991 p 580) has indicated that inter-departure time data measured from aggregate traffic can not accurately reflect true inter-arrival time. Smith and Kain (1991) pointed out that at network load level of 25%, there are 55% of all transmitted packets that incurred some queuing delay. The delay could be in the form of: the queuing delay at Ethernet interface, the backoff delay due to collisions or the delay due to deference mechanism.

The bus length also affects the packet transmission delay as it determines the propagation delay parameter. But, this delay is external delay rather than internal delay as described above. This kind of delay has more influence on measured packet time stamps due to relative location of the host that functions as a recorder station. Smith and Kain (1991) have also indicated that aggregate traffic data often does not account for the network propagation delay. However, propagation delay may affect packet transmission characteristic of an individual station. This is particularly true for a long bus. The high delay experienced by an acknowledgement packet will in turn affect packet transmission characteristic of a station waiting for acknowledgement.

An other network parameter that may affect the packet transmission characteristic of individual station is distribution of station locations on the bus. In their simulation study of Ethernet network, Gonsalves and Tobagi (1988) pointed out that individual station performance varies with the location of the station.

### 2.4.2 User Parameters

User parameters refer to the individual traffic characteristics generated by an individual station. This can be represented by station's activity level, i.e. rate at which a station offers its packet to the Ethernet protocol to be transmitted. This metric (in term of packets/second) can be estimated by measuring the packet departure rate of a station under consideration, that is by dividing the total number of transmitted packets to the duration of measurement. The station activity level is related to its packet generation processes (i.e., from Ethernet client layer) through its mean packet arrival rate (packet arrival processes). Generally, the higher the station arrival rate, the higher its packet departure rate. It is also related to the protocol parameter known as flow control and its related buffering mechanism (Papadopoulos 1992, pp. 80-83). This protocol parameter will be discussed shortly.

Individual traffic characteristic can also be represented by different applications running in stations. For example, file transfer applications may have different packet size distribution compared to word-processing or telnet applications (Morino & Takahara 1992). Thus, different user characteristics may be represented by different packet size distributions and their appropriate inter-arrival time distributions. Since longer packet will have different inter-departure time compared to the shorter one, this user characteristic will obviously affect packet transmission characteristic of individual stations.

### 2.4.3 Protocol Parameters

We consider here the buffer size at each station as a protocol parameter that may affect the null hypothesis. This buffer size is closely related to the performance of an individual station through the window flow control mechanism. Papadopoulos and Parulkar (1993) pointed out that throughput of individual stations in Ethernet network may be improved by increasing the buffer size at each station. Improving throughput means increasing packet transmission rate of individual station. For a fixed packet size, higher transmission rate simply means shorter packet inter-departure time.

Previously, Gonsalves and Tobagi (1988) have also studied the effect of buffer size on the Ethernet protocol. They concluded that increasing buffer size will reduce the station idle time, which in turn will increase the station throughput.

Other protocol parameter that may affect the null hypothesis is the backoff algorithm used by Ethernet to handle collision events. In the Ethernet protocol this algorithm is known as *truncated binary exponential backoff* (Wolfberg 1983 p 51). Gonsalves and Tobagi (1988) indicated that in high collision rates this algorithm should be modified to improve station performance. Other studies have also suggested modification of Ethernet backoff algorithm to account for phenomena known as *capture effect* (Hayes & Molle 1995; Ramakrishnan & Yang 1994). This is an unfairness condition in which a collided station is more likely to collide again thus increasing packet delay.

## 2.5 Problems in LAN Traffic Measurement

Other problems related to the measurement-based traffic modelling is the accuracy of the measurement system. Perhaps the most widely used tool in LAN traffic measurement is *tcpdump* (Jacobson, Leres & McCanne 1994). Tcpdump is a TCP/IP protocol analyser that can capture packet headers from Ethernet, print information from the header as well as packet's time stamp. In SunOS, tcpdump is known as *etherfind* (Hunt 1994 p. 262). This tool has been used in previous studies (Cinotti et al. 1995; Deng, Bugos & Hill 1996; Robert & Lambrecht 1995; Ryu 1996). Other monitoring tool known as *snoop* (snoop 1992) which essentially has about the same capability with tcpdump has also been used in (Liu, Anido & Chicharo 1994) (pers. comm. with Liu, September 1995).

In traffic modelling, perhaps the most important information resulted from traffic monitoring tools are the packet's time stamp, packet size and source-destination address. From these three variables, several important statistics can be obtained, such as packet inter-departure time distributions, time series data, packet size distributions and traffic source-destination patterns. The most critical one is the packet time stamp data, while other variables are normally independent of measurement system accuracy. Tcpdump's manual (Jacobson, Leres & McCanne 1994) mentioned that the tcpdump's time stamp accuracy depends on the kernel's clock (e.g., 10 ms on Sun-3). A previous study (Kao 1993) revealed that there are at least 3 factors that may affect the results of traffic monitoring system such as tcpdump or snoop. They include buffer size, CPU speed and memory space.

We have also detected the poor accuracy of packet's time stamp data from tcpdump and snoop. We performed a simple test on Ethernet traffic data collected using tcpdump and snoop running on SPARC station 5. We collected several data sets containing various packets, ranging from 10,000 packets to 1,000,000 packets (see Table 2.1). The data contains only non-error packet information, i.e., packets that have been transmitted successfully. Network utilisation during data collection were mostly very low. Given the bus rate of 10 MBps, packet's time stamp and packet size, we calculate the packet duration time and the time difference between successive packets on the bus. We found that overlapping packets (see Figure 2.2) are present in each data set. This overlapping phenomenon was also detected even in individual data traffic that is obtained by extracting source address in packet header using tcpdump or snoop. It is unlikely that packets from the same source will collide with each other. Since the data does not contain collided packets, the possible source of this error could be the rate of the bus which may be higher than 10 MBps (it is most unlikely) or simply because the time stamp is not accurate due to some factors as pointed out by Kao (1993). One of tcpdump's authors (Leres 1995, pers. comm., 27 October) and Papadopoulos and Parulkar (1993) suggested to use special dedicated hardware to obtain more accurate time stamp. Table 2.1 presents the statistics of testing the overlapping events. The overlapping intensity is shown as a ratio of number of overlapping events to the total packet.

Surprisingly, using that simple test, we also have detected the same error in the data set that had been claimed was collected using a high time-resolution special dedicated hardware in (Leland & Wilson 1991) (see Table 2.1). Part of this data (collected in 1989) was available in ITA (The Internet Traffic Archive, URL: http://www.town.hall/Archives/pub/ITA) site and also in Bellcore site (ftp.bellcore.com under the directory /pub/wel/lan-traffic). This data is known as *Bellcore Data* (Duffield et al. 1994) and has been used as an object of study by other work (Duffield et al. 1994; Robert & Boudec 1996; Ryu & Lowen 1996). In fact, the well-known aggregate traffic model 'Self-Similar Traffic' originally proposed in Leland et al. (1993) was a result of this kind of data set. Perhaps another research is required to investigate the effect of this error in time series data and on the model of self-similar traffic before this error can be justified.

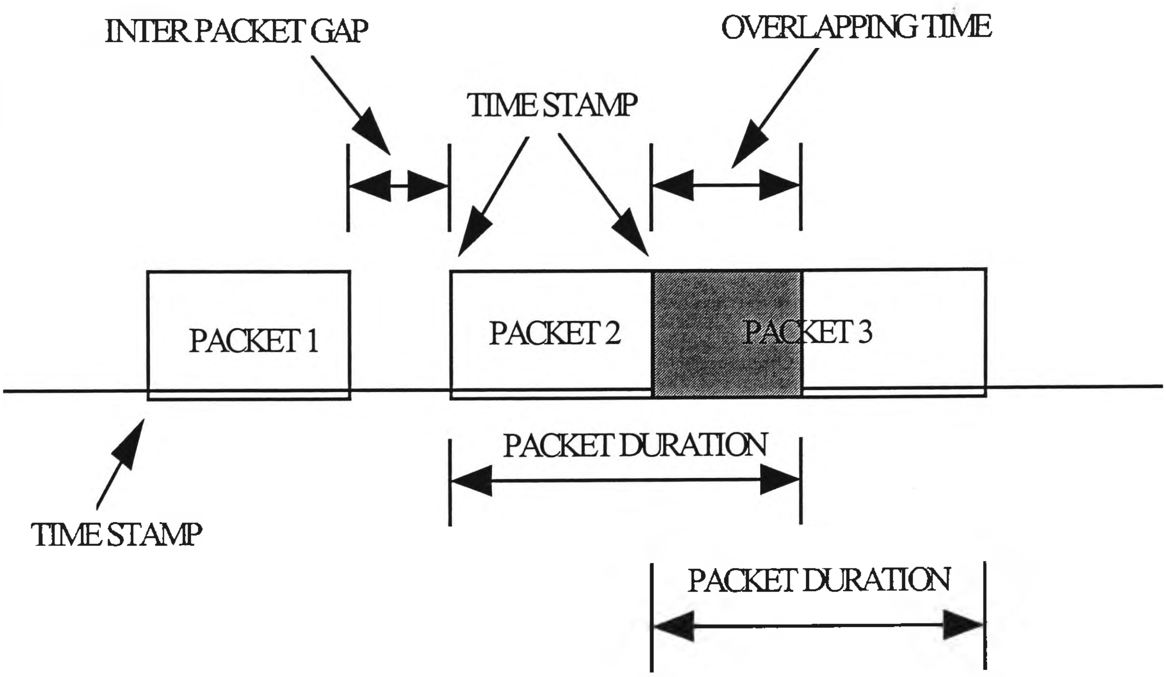


Figure 2.2 Overlapping Packets

In conclusion, the traffic monitoring system that is known as Spying Network Monitor (Smith & Kain 1991) is not a good tool to study packet inter-arrival time and to model traffic source. This aggregate traffic monitoring system can only record the time when packet is seen in the network while accurate packet inter-arrival time can only be measured within a station. However, this monitoring system is useful for network management purposes since this later task does not require accurate inter-arrival time statistics (Smith & Kain 1991).

Table 2.1 Statistics of Overlapping Packets

| No. | Data Set | Total Packet | Utilisation | Maximum Overlapping Time | Percentage of Overlapping |
|---|---|---|---|---|---|
| 1. | tcpdump 1 | 10,000 | 0.801 % | 1.076 ms | 6.248 % |
| 2. | tcpdump 2 | 1,000,000 | 1.970 % | 1.199 ms | 8.295 % |
| 3. | tcpdump 3 | 1,000,000 | 5.150 % | 1.197 ms | 10.921 % |
| 4. | snoop 1 | 10,000 | 2.104 % | 1.195 ms | 8.725 % |
| 5. | snoop 2 | 85,000 | 7.872 % | 1.201 ms | 10.341 % |
| 6. | Bellcore 1989 pAug.TL | 1,000,000 | 11.142 % | 1.196 ms | 10.071 % |
| 7. | Bellcore 1989 pOct.TL | 1,000,000 | 4.647 % | 1.172 ms | 15.242 % |

Given the above fact and limited resources available, it is still useful to study the validity and the sensitivity of the null hypothesis. If we know the applicability range (i.e., within which range of network utilisation) of this hypothesis, then we may still utilise an existing network monitoring tool to conduct the traffic modelling study.

## 2.6 Simulation as an Evaluation Technique

There are three techniques commonly used in evaluation study (Jain 1991 pp. 30-33). They include analytical modelling, simulation and measurement. Jain (1991) also discusses several criteria to select the most appropriate evaluation technique for a given problem. For the purpose of this project, we have selected the simulation technique due to the accuracy and cost reasons.

Compared to simulation techniques, for a given problem, analytical techniques generally tend to make more simplifying assumptions for tractability. This results in lower accuracy for analytical results. Previous studies of Ethernet networks have used simulation to obtain more accurate results. These involved studies conducted in (Al-Salqan et al. 1991; Devai, Kerrigan & Molloy 1990; Gonsalves & Tobagi 1988; Prasad & Patel 1988). Furthermore, in this study we need more freedom to experiment with multiple parameters. The requirements are supported well by simulation because it can incorporate more details but requires less assumption.

Compared to measurement techniques, simulation results may be less accurate. But, given the above fact about traffic measurement problems, the accuracy of measurement study may be questionable. One possible way to obtain more accurate packet inter-arrival time measured data is to use internal station monitors as suggested by Smith and Kain (1991). But, this also means more investments are required on the protocol hardware and software. Furthermore, any modification on protocol software to include traffic probing capability will result in disturbance of actual packet inter-arrival time. This is due to additional overhead introduced by traffic probing code (Papadopoulos 1992, p 70).

However, simulations can not stand alone and require some form of validation. Rules of validation (Jain 1991, p 32) point out that the results of simulation model can only be justified if they have been validated by analytical or measurement results. This task is probably the most difficult part in simulation technique (Sadiku

& Ilyas 1995, p 45). However, several techniques are available for the validation purposes (Devai, Kerrigan & Molloy 1990; Jain 1991, p 420). This subject will be explored and employed later in Chapter 3.

## 2.7 Related Analytical Studies and Self-Similarity Issues in Traffic Modelling

It is worth to noting that there exist several analytical studies related to the study in this thesis. Started with the elegant paper of Tobagi (1982) and followed by (Takahashi, Matsumoto & Hasegawa 1986; Matsumoto, Takahashi & Hasegawa 1990a; 1990b; Tan & Tsai 1996), these studies actually obtained the probability expression of packet inter-departure time of aggregate traffic by assuming Poisson process of individual packet arrival. All of these studies used simplifying assumption of Poisson process for packet arrival time to make their analysis tractable. However, several recent observations of self-similar behaviour in network traffic have argued that Poisson assumptions may be useful only in analysis but will not always agree with real word phenomenon (Paxson & Floyd 1994). In this study, we are going to use second order statistic tests to test the presence of self-similarity or fractal behaviour in aggregate traffic simulation data. These test's results will be useful as a reference when comparing the null hypothesis test results of simulation data generated by different traffic models. We will describe the second order statistic tests more detail in Chapter 4. For more information about self-similarity issues in traffic modelling we refer the reader to a bibliographical guide by Willinger, Taqqu and Erramilli (1996) and to (Ryu & Lowen 1996).

# 3. Simulation Model

## 3.1 Outline

We have argued in the previous chapter that the method of simulation is the most appropriate tool in this thesis. This chapter is intended to outline the development of an Ethernet network simulation model. The Ethernet network has been considered as an object of this study because most of the measurement-based traffic modelling studies, described in Chapter 2, used the traffic data collected from Ethernet networks.

Section 3.2 describes the event-based simulation approach used to develop the model. The model is constructed to imitate behaviour of real Ethernet network. In doing so, several underlying assumptions as well as previous supporting studies are utilised. The term frame and packet will be used interchangeably. The discussion covers only transmitting site of an Ethernet network since this study is concerned with traffic stream from Ethernet's client layer down to the bus. We use the fact of asymmetric traffic loads (Senior, Rehal & Wiseman 1992) to focus our study at the transmitting site of Ethernet network. As also observed in our departmental LAN traffic characteristics, more than 90% of traffic data is transmitted from server to its client which means that the transmitting site of Ethernet network is more dominant in describing LAN traffic characteristics.

To facilitate model construction, a queuing model abstraction is introduced. In addition, to minimise external factors that may affect the null hypothesis test results on interested parameters as described in Chapter 2, several external factors are carefully localised while maintaining the accuracy of simulation model. This means the model will only incorporate those inherent factors of the Ethernet protocol while other external factors that may affect the null hypothesis are minimised. The purpose of this effort is to make sure that the null hypothesis test result is the ones with affecting parameters under consideration. Model's features and limitations are summarised at the end of Section 3.2.

Before the simulation model can be used to test the null hypothesis, it is necessary to verify and validate the model. These techniques are applied to the simulation model in Section 3.3. Previous analytical and measurement studies are used to validate the model. In addition, to manage simulation data collection, Section 3.4 describes transient removal and stopping criteria techniques.

## 3.2 Ethernet Simulation Model

### 3.2.1 Discrete Event-Driven Simulator

The three most widely used software simulation models, as opposed to hardware simulation model or emulation, (Jain 1991, pp. 403-408) are the Monte Carlo

Simulation, Trace-Driven Simulation and Discrete-Event Simulation (DES). Since processes in an Ethernet network involve statistics that change over time, the Monte Carlo Simulation is not appropriate to model the Ethernet network. Trace-driven Simulation could be relevant to our study. Unfortunately we do not have an accurate time-ordered record of events on a real system to drive simulation. Moreover trace-driven assumes the trace is independent of the network state. Therefore we consider the Discrete-Event Simulation is the most suitable model in our study. In fact, previous Ethernet simulation studies also used the DES technique (Al-Salqan et al. 1991; Devai, Kerrigan & Molloy 1990 Gonsalves & Tobagi 1988; Prasad & Patel 1988;). This simulation technique is also employed extensively in (Sadiku & Ilyas 1995). Typical flow design of DES (Frost & Melamed 1994) is depicted in Figure 3.1.

A list of events needs to be determined in advance. An algorithm is then used to scan one event at a time to be processed, based on event's time-parameter, which is normally related to the simulation clock. The simulation clock is advanced every time an event is processed. The program will update dynamically the event list prior to scan the next event. More detail explanations of DES models can be found in many textbooks such as (Fishman 1978; Jain 1991; Law & Kelton 1991).

It is worth noting that there exists a ready to use Ethernet simulation packet such as OPNET (OPNET 1993). Unfortunately, we found that this package is not user-friendly enough to use in our study. We encountered several difficulties when trying to use OPNET in our study including problems in incorporating user-specified codes for certain traffic source models. Furthermore, to apply our required statistical analysis, a specific format for the simulation data is required and it is not easy to obtain this formatted data in OPNET. All of these have motivated us to develop our own simulation model rather than using the existing package.

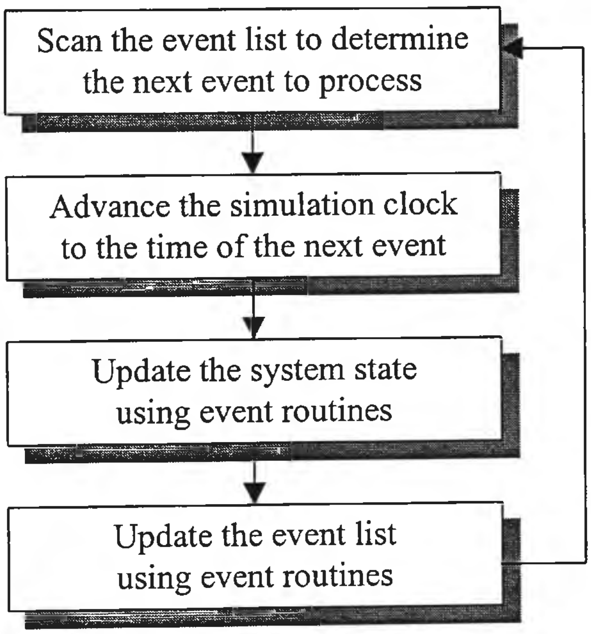


Figure 3.1 Discrete Event Simulation Flow Design

### 3.2.2 Ethernet Protocol Overview

The Ethernet network is a bus type local are network (LAN) which uses a random-access protocol, as opposed to a controlled-access protocols, known as CSMA/CD (Carrier Sense Multiple Access with Collision Detection). The term multiple access implies a mechanism in which all stations on the bus can receive the transmitted frame at the same time, including the transmitting one (Hsu 1996, p 198). This medium access control (MAC) protocol, also known as IEEE 802.3, is intended to regulate and provide efficient and fair access of the transmission media. The CSMA/CD protocol can be related to the first two layers of the OSI Model as depicted in Figure 1.1. Figure 3.2 presents the position of the Ethernet MAC protocol in peer-to-peer layered communications architecture (Archie et al. 1993).

In Figure 3.2, the logical link control layer is assumed to be part of the Ethernet MAC layer. The higher layers are any network protocols, known as client layers in Ethernet Standard terminology (Wolfberg 1983 p 111), that use the service provided by Ethernet data link (e.g., Internet Protocol IP). The client layer sends a packet down to the Ethernet MAC layer. If the later is idle, then a packet is scheduled for transmission, otherwise the packet is queued. This queuing mechanism is an important part of our simulation model and will be explained later in the following section.

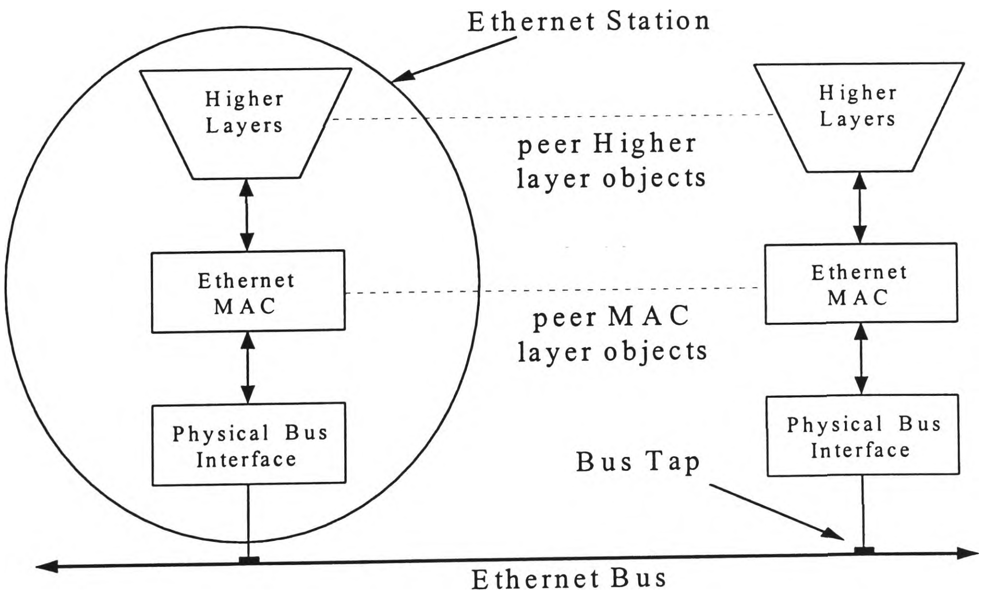


Figure 3.2 Position of the Ethernet MAC in Layered Communications Architecture

### 3.2.2.1 Carrier Sensing

A station having a packet ready to transmit has to 'listen' to the bus, through its physical bus interface, before attempting to transmit its packet. This mechanism is known as *sensing*. It continues doing so and withholding transmission until after no carrier is detected on the bus. This process is called *deference.* In standard Ethernet networks, as soon as a station detects a free bus (no carrier is present), it may transmit its outstanding packet after a constant short period of deferring has elapsed. This additional constant delay is called *interframe spacing* (9.6 μs) and it is intended to provide inter-packet recovery time for other data link controllers and for the physical channel (Wolfberg 1983, p 50). This ensures that all stations can distinguish the end of current packet and the beginning of the next packet on the bus. The immediate action of packet transmitting after carrier sensing is due to technique adopted by Standard Ethernet and IEEE 802.3 called *1-persistent.* It means, as soon as the bus is free, a ready station may transmit its packet with probability 1 (Boggs, Mogul & Kent 1995).

The carrier sensing process and its related deference period is one mechanism used by Ethernet's MAC protocol to deal with collision events. Another mechanism known as *Truncated Binary Exponential Backoff* will be described shortly. The effectiveness of this deference mechanism depends on the bus physical factor known as *signal propagation delay,* i.e., the time required by the signal to travel through a given bus media and be detected by each station on the bus.

Given that each station only transmits when it senses the bus as free, there is still a possibility that two or more stations will transmit at approximately the same time. When this happens, the signals collide and distort each other. The packets can not be received correctly and packet re-transmission is required. This is because a packet transmission attempt at each station is made only on basis of local information, i.e. from its physical bus interface, not from overall situation of the bus.

Given that deference mechanism is intended to avoid collision event, the failure of this mechanism is usually not the causal factor for collisions, because the propagation delays are typically quite small (Archie et al. 1993). This is another important factor to be considered in developing simulation model. In fact, the collisions are most likely due to the built in fairness mechanism of Ethernet MAC. If more than 1 station has been deferring and waiting for the interframe gap to elapse, then, when the bus is detected free, those stations would assume falsely that the bus is free (locally in fact) and begin to transmit their packet at approximately the same time. In this situation, their packets will collide with each other. This situation is unavoidable in Ethernet networks and normally happens within one period of propagation delay following the packet transmission. This critical period is known as the *vulnerable period* (Gonsalves & Tobagi 1988) or *collision window* (Black 1993, p 128). Given the propagation delay between two attempting station (say **j** and **k**) is $\tau$ seconds and the time it takes a station to detect collision is $\xi$

seconds; then the vulnerable period of station **j** starting transmission at time **t** is the interval

$$[t - (\tau + \xi),\ t + (\tau + \xi)] \qquad \textbf{(Fml 3.1)}$$

A packet from station **k** will collide with **j**'s packet only if **j** starts transmission in that interval. Other stations located closer to station **j** would certainly have a shorter interval about **t** during which their packets will collide with **j**'s transmission.

### 3.2.2.2 Collision Handling

When collision occurs, the transmitting stations will detect a difference between the signal being transmitted onto the bus and the signal being received. The transmitting stations continue to transmit an extra encoded bit sequence (at least 32 bits but not more than 48 bits) used for *collision enforcement*. This mechanism is called *jamming* and is intended to ensure that duration of the collision is long enough to be detected by all other transmitting stations involved in the collision. The time that has to elapse before the bus is sense idle in collision event depends on three variables: the one way propagation delay between two collided station, $\tau$, the time it takes a station's device to detect interference, $\xi$, and the jamming bit duration, $\zeta$. Tobagi and Hunt (1987, pp. 320-321) pointed out that the time until the bus is sensed free by all stations is

$$3\tau + \zeta + \xi. \qquad \textbf{(Fml 3.2)}$$

It is worth to noting that variables $\tau$, $\xi$, $\zeta$ in Formula 3.1 and 3.2 are implemented in simulation program by assigning their maximum values. The propagation delay between stations is set the same due to the star configuration used in simulation model.

#### 3.2.2.3 Truncated Binary Exponential Back-off Algorithm

If the collision period has elapsed, the stations having packets to transmit can be grouped into two groups. The first group is those stations that were involved in the previous collision event. For this group, each station has to voluntarily delay its transmission according to the algorithm called *Truncated Binary Exponential Back-off*. The random time delay resulting from this algorithm is an integral multiple of a slot time which is defined to be 512 bit times or 51.2 μs in 10 MBps network. The total amount of delay (in unit of slot times) before $\mathbf{n}^{\text{th}}$ retransmission attempts is defined to be a uniformly distributed random integer $\mathbf{r}$ in the range

$$\mathbf{0 \leq r \leq 2^{k} \qquad \text{where } k = \min(n,10)} \qquad \textbf{(Fml 3.3)}$$

If the first 10 attempts were not successful, a station may continue to try until maximum attempts of 16 with the value of k bounded to 10. If all 16 attempts are unsuccessful, this event is reported as an error to the client layer of Ethernet MAC. This usually happens in a very high loaded network. The current packet is normally discarded. On the other hand, once a packet is transmitted successfully, the station reset its counter back to zero in preparation for the next packet.

The second group is those stations that were not involved in the previous collision event. For this group, they may transmit their packets immediately (after 9.6 μs delay). They do not have to delay their transmission according to Ethernet's backoff algorithm, because the Ethernet protocol employees only *transmitter-based collision detection scheme* (Shoch et al. 1982).

### 3.2.3 Packet Flow and Queuing Model

In order to develop a good model, it is necessary to understand how a frame is moving from the client layer to the Ethernet MAC before it is transmitted. Closely related to this process is a queuing abstraction along the data path. Since this study is concerned with the effect of data link layer (Ethernet MAC layer) protocol to the packet stream, the discussion is limited to one layer above the Ethernet MAC layer, e.g., TCP/IP layer. In this context, the TCP/IP layer is considered as a traffic source generator that produces a packet stream. This stream is then processed by the Ethernet MAC layer before transmission to the bus. Detail explanation of message flow process from application layer to the physical layer in a typical implementation of TCP/IP protocol on Ethernet network can be found in (Huang & Chen 1992; Leffler et al. 1989, pp. 279-390; Papadopoulos 1992, pp. 47-78; Stevens & Wright 1995, pp. 18-20).

In their study of the performance of SunOS Inter-Process Communication (IPC), Papadopoulos and Parulkar (1993) have identified several queuing points in different layers. Their study was based on a typical implementation of TCP/IP protocol using UNIX 4.3 BSD operating system running on SPARCstation 1 with AMD Am7990 LANCE Ethernet controller. This study indicated that processing rate of TCP/IP protocols is estimated at about 22 MBps while processing rate of Ethernet driver is about 204.8 MBps (based on 40 μs delay to process 1024 bytes data). Given that the Ethernet bus rate is 10 MBps, the above figures clearly indicate the presence of a queuing point at the network interface layer as depicted in Figure 3.3. Since the Ethernet bus is a serial medium and the Ethernet protocol only processes one packet at a time, this queue can be considered as a first in first out (FIFO) queue (Taylor, Oster & Green 1983, pp. 733-739). The first packet on the queue will be held until it is transmitted successfully or discarded due to excessive collisions.

The network interface layer is a layer where the interaction between the TCP/IP protocol software and a network hardware device driver (e.g., Ethernet card) takes place (Papadopoulos 1992, p 49). Depending on its design, the Ethernet driver itself could have additional buffers to queue packets ready for transmission (Jeffrey C. Mogul 1997, pers. comm., 7 January; Stevens & Wright 1995, p 112). For the purpose of our study, these 2 queues will be considered as 1 queue, i.e., the queue of packets before they are processed by the Ethernet MAC. This assumption is based on typical processing speed of three layers under consideration: TCP/IP protocols, Ethernet MAC layer, and physical layer or bus rate as discussed above.

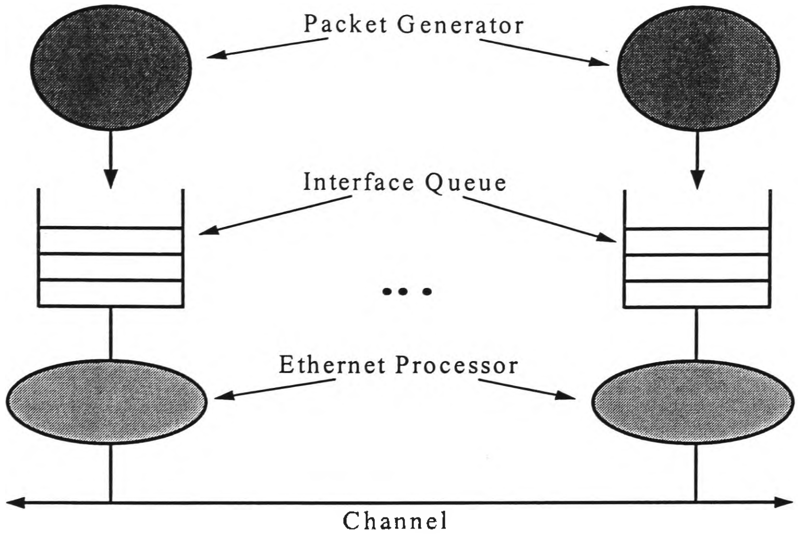


Figure 3.3 Queuing Abstraction in Ethernet Network

Papadopoulos (1992, p 122) indicated that in its default setting, the interface layer queue size can be as much as 50 packets. This size of buffer guarantees no packet loss due to overflow for 12 simultaneous connections with maximum window of 4 (4096 bytes). This implementation will allow at most four Ethernet packets to be transmitted with every window. In light of flow control mechanism, this is equal to the four-packet stop-and-wait protocol. The station can only send at most four packets and has to wait for an acknowledgement before sending the next packet.

The size of packet queue or buffer will be one of our test parameters. In a real Ethernet driver, the buffer memory area is a shared area between the transmit and receive areas (Wolfberg 1983, p 447). The ratio between transmit and receive areas is programmable. In most systems, typically 2 full-length packet areas are allocated for the transmit queue. However, in the previous Ethernet simulation studies

(Mazraani & Parulkar 1992; Smith & Kain 1991) this queue size is considered very big (infinite queue).

Mazraani and Parulkar (1992) used a buffer of infinite capacity to avoid packet loss due to buffer overflow. Based on their measurement, Smith and Kain (1991) considered that the buffer space limitations are not a significant problem. They thus assigned a big enough size of packet buffer to their simulation model because the buffer limits in the stations they monitored was never reached. This is actually the result of real flow-control mechanism between packet generator and buffer, which is norm in most applications (Boggs, Mogul & Kent 1995). Since the packet arrival processes of flow-controlled applications is much harder to model, most simulation studies adopt an assumption of infinite packet buffer to avoid modelling the packet arrival time distribution of the flow-controlled mechanism. However, as pointed out by Stevens and Wright (1995, p 111), the packet discarding processes do happen when a driver's transmit buffer is currently full.

### 3.2.4 Modelling Assumptions

As mentioned in Chapter 2, there are at least 3 parameters that may affect the null hypothesis. Since this thesis is concerned with the influence of data link protocol on traffic stream, it is necessary to isolate the affecting factors. The rationale behind this is to obtain the test results concerning the affecting factors of interest.

One of our test parameters is a station's activity level, which will be modelled by a station's mean packet arrival rate. More detail about test parameters in this study will be described in Section 5.2. In traffic measurement, a station's activity level can be measured approximately as the packet departure rate in term of packets/second. This statistic is closely related to station throughput. As pointed out by Gonsalves and Tobagi (1988), this statistic is affected by a station's location on the bus. Therefore, to obtain location-independent test results for station's activity level parameter, it is necessary to use a fair network topology in the Ethernet simulation model. In fact, this topology is the pessimistic one that is usually used in analytical studies of CSMA/CD networks (Lam 1980; Matsumoto, Takahashi & Hasegawa 1990a; 1990b; Takahashi, Matsumoto & Hasegawa 1986; Tan & Tsai 1996; Tobagi 1982; Tobagi & Hunt 1987, pp. 318-339). This configuration belongs to a non-linear topology (Gonsalves & Tobagi 1988) known as *star network.* It is depicted in Figure 3.4.

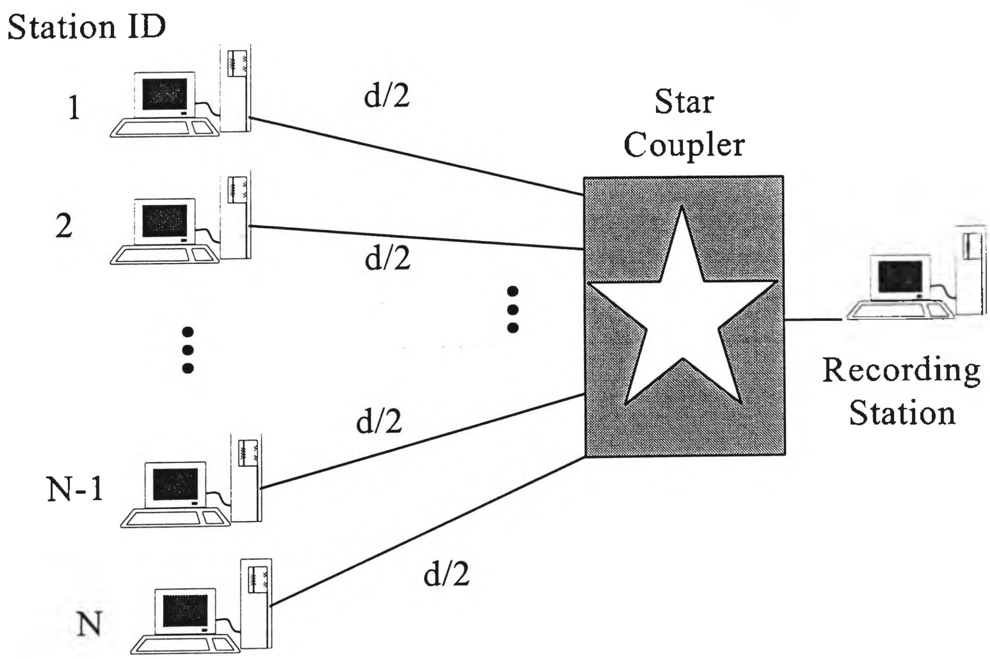


Figure 3.4 Ethernet Star Network

For the purpose of this study, there are several advantages raised by using a star network configuration. First, any dependency on station location (mainly location-dependent parameters) is relaxed. Second, while analytic models of Ethernet network are not yet available (Gonsalves 1987, p 393), for all circumstances, this star configuration also facilitates model validation to analytical models (CSMA/CD models). Third, as depicted in Figure 3.4, we may assume the centre location as the place where a traffic recorder is located; thus there is no need to adjust a packet's time-stamp due to relative position between the traffic recorder and transmitting stations. Forth, since propagation delay between any station is the same, this will reduce complexity in station's carrier sensing calculation that is imperative in case of a linear bus topology.

Another modelling assumption deals with the bus length. Most of our experiments will consider a short bus medium. Thus, the simulation model will not consider packet repeaters or gateways. However, as one of network parameters, we will investigate the sensitivity of the null hypothesis against this parameter by varying the bus length to incorporate the effect of high propagation delay.

Since the important statistic for testing the null hypothesis is the packet inter-arrival time distribution, we need to minimise other factors that could distort this distribution data in our simulation. We thus assign a very big packet buffer size at each station. By doing so, any distortion of the packet inter-arrival time distribution due to packet discarding process when queue is full will be minimised. However, we will also test the influence of packet buffer size limitation on the sensitivity of the null hypothesis.

### 3.2.5 Model Development

Given the background discussion above and after reviewing previous simulation studies of Ethernet and CSMA/CD networks (Al-Salqan et al. 1991; Devai, Kerrigan & Molloy 1990, Gonsalves & Tobagi 1988; Mazraani & Parulkar 1992; Prasad & Patel 1988; Rives et al. 1988; Sadiku & Ilyas 1995), a discrete-event simulation model of an Ethernet network was developed based on four main events as proposed by Sadiku and Ilyas (1995, pp. 107-133). However, to obtain more accurate simulation results, we have improved their CSMA/CD model to incorporate standard Ethernet protocol features. Main improvements from their CSMA/CD model are: relaxing slotted time assumption, incorporating collision detection mechanism based on the vulnerable period calculation (Fml. 3.1), adding efficient carrier sensing mechanism based on star configuration advantage thus speeding up simulation execution, including standard truncated binary exponential backoff algorithm based on 512 bit times unit and incorporating variable standard packet sizes. Those improvements result in a good agreement between simulation results and measurement results (see Section 3.3.2 of validation result). However, their CSMA/CD simulation model does have some interesting and important aspects. These include the use of packet identification number (ID) and Event 2 of their model. We found that packet ID is very helpful in statistical calculation associated with a packet. In addition, the second event of their model, packet transmission attempt, even though it appears to be redundant for our analysis, it was found to be useful for speeding up the simulation execution time.

In their CSMA/CD model, Sadiku and Ilyas used four main events (packet arrival, transmission attempt, collision check and packet departure). The second event in their model (i.e., transmission attempt) appears to be redundant and may be omitted without affecting the final simulation results. This event is included to account for an unfair process in which a scanning module of DES program will always pick up a station represented by the smallest number (Sadiku & Ilyas 1995, p. 121). In fact, given the star configuration, any station attempting to transmit within the same vulnerable period will obviously be involved in collision. Thus, the order of stations is not relevant and from this point of view the second event is redundant. However, when this second event is omitted in our test model, the simulation execution time is increased greatly. This is due to heavy load on the middle event (combination of transmission attempt and collision check) creating a number of loops and consuming additional execution time. Thus, to maintain simulation speed, we designed our simulation model with four main events as in its original model, although the second event does not produce any desired statistics.

We also have relaxed the slotted operation assumption used in Sadiku and Ilyas's model. Slotted operation basically assumes the presence of a master slot-time clock (Boggs, Mogul & Kent 1995). Each transmission is synchronised to start only at the beginning of time slot. Sadiku and Ilyas used this assumption to simplify the collision detection mechanism. In their model, a collision is defined when two or more stations attempt to transmit within the same time slot. Therefore, one propagation delay has been used as a unit time in slotted operation. However, a previous analytical study (Takahashi, Matsumoto & Hasegawa 1986) has pointed out that unnatural discretisation of the time scale tends to underestimate the delay as traffic increases.

Owing to star configuration and employing Formula 3.1, we have relaxed this time slotted assumption in our simulation model. In addition, by incorporating other features of Standard Ethernet, our model can be considered as an improvement of model proposed by Sadiku and Ilyas (1995).

### 3.2.6 Event Modelling and Statistic Probes

An event is a program that is executed at discrete simulation times. The four main events in our model are: Packet Arrival, Channel Seizure Attempt, Collision Check and Packet Departure. Within main events there exist some sub-events deal with mechanisms of Ethernet MAC as discussed above. Figure 3.5 illustrates the relations between the four main events and conditions that activate each event. Based on these relations our simulation program is developed with the flow chart as depicted in Figure 3.6.

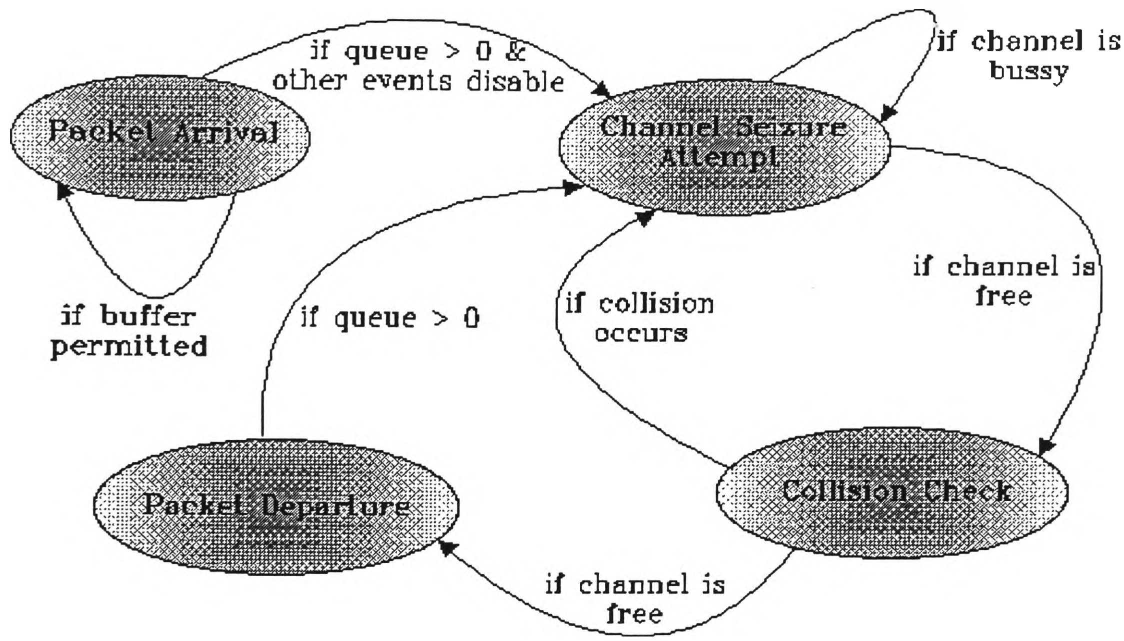


Figure 3.5 State Transition Diagram of Ethernet Simulation Model

For the purposes of analysis, statistical data collection is done by placing statistic probes at appropriate point in particular events. The following subsections describe the four events and statistic probes as well as the measured statistics.

### 3.2.6.1 Event 1: Packet Arrival

This is the only event that has to be initialised when the simulation program is started. Other events are initially assigned a very large value (called the DISABLE value) to indicate that they are not active. However, during the simulation run, there may be maximum two events being active at the same time. Event 2 (Channel Seizure Attempt), Event 3 (Collision Check) and Event 4 (Packet Departure) can not be active at the same time. At any time, there is only one packet attempt to be transmitted by any station. Thus, Event 2, Event 3 and Event 4 may active in turn. Event 1 may coexist with other events at any time.

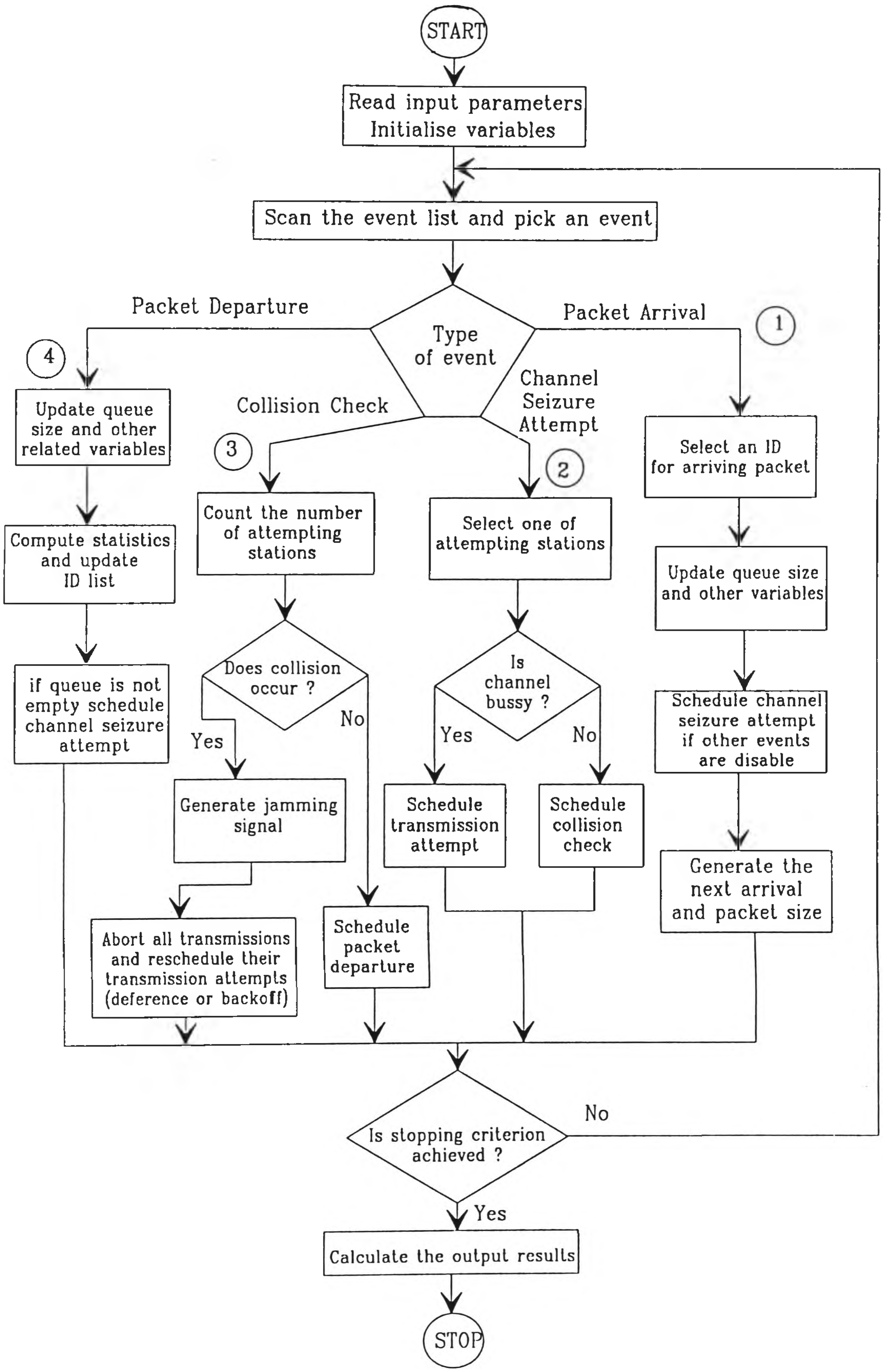


Figure 3.6 Flow Chart of Ethernet Simulation Program

The packet arrival event deals with the generator module that generates packets based on specified inter-arrival distribution. This generator actually models the traffic stream from higher layer (e.g., TCP/IP). The mean packet arrival rate, as a measure of station activity level, is the mean of inter-arrival time distribution. As in the algorithm used by Sadiku and Ilyas (1995, p112), each arriving packet brings its unique identification number (ID). The size of this ID equals to the number of station on the bus multiplied by buffer size or queue size (in number of packets) at each station. This unique ID is very useful in all calculations related to a packet such as packet arrival time, packet delay, packet departure time or packet size. The program always maintains an array of these IDs. Successfully transmitted packets will release their IDs to be used by newly arriving packets, enabling the simulation program to deal with a large number of packets.

Each arriving packet also has its own size, which is generated based on a predefined packet size distribution. The packet size distributions (see Section 5.3 and 6.3) are obtained either from our own measurement data collected using TCPDUMP or from previous measurement studies (Gusella 1990; Khalil, Luc & Wilson 1990; Morino & Takahara 1992; Robert & Lambrecht 1995). To simplify packet generation processes, it is usually assumed independent of inter-arrival time and packet size (Frost & Melamed 1994). The generated packets are then placed inn FIFO queue before being processed by the Ethernet MAC.

The following statistics are collected on the packet arrival event :

- time stamp of the packet's arrival and it's inter-arrival time
- packet size
- starting time of the packet, i.e., the time of the first attempt to transmit the packet (beginning of Event 2 for each packet).

- Offered load of each station, which is defined as the number of useful bits per second that the station attempts to transmit on the channel (Mazraani & Parulkar 1992). For simplicity, header and trailer are included as useful bits. Packet preamble is not considered as useful bits in offered load calculation. Figure 3.7 depicts Ethernet Frame format. The valid frame size is from 64 octets to 1518 octets.

| Preamble | Destination Address | Source Address | Type | DATA | Frame Check Sequence |
|---|---|---|---|---|---|
| 8 bytes | 6 bytes | 6 bytes | 2 bytes | 46 to 1500 bytes | 4 bytes |

Figure 3.7 Ethernet Frame Format (Wolfberg 1983, p 46)

The starting time is used to calculate transmission delay of each packet which is defined as the elapsed time between its first attempt and its successful transmission (include packet duration) (Boggs, Mogul & Kent 1995). Offered load of each station is used to calculate total offered load at the end of simulation time. Total offered load is defined as the sum of all individual offered loads normalised to 10

MBps or in percentage of 10 MBps. In the initialisation, the theoretical total offered load is calculated as traffic intensity $\rho$ given by (Sadiku & Ilyas 1995, p 117)

$$\rho = \mathbf{AR * N * PL * 8 / BR} \qquad \textbf{(Fml 3.4)}$$

where AR, N, PL, BR are mean station's arrival rate , number of station, mean packet length and bus rate respectively. The factor 8 is for byte to bit conversion.

The most important statistic collected from this first event is the packet inter-arrival time distribution. This distribution will be tested against packet inter-departure time distribution generated from the packet departure event. The statistics of this test will determine how sensitive the null hypothesis is. The statistical methods used to address this test will be described in Chapter 4.

Other processes happening in this event include: Event 2 scheduling, station's queue maintenance and next arrival scheduling. As illustrated n Figure 3.5, this first event can only be triggered by the arrival of next packet. If the queue is full, the queue full indicator is set and the next arrival packet is discarded. It is in line with real Ethernet driver operation (Stevens & Wright 1995, p 111). In our simulation program, this discarding process simply means the arriving packet will not be processed by packet arrival processor (see Figure 3.6). The queue full indicator is also useful to prevent the simulation program from quitting abnormally since programming language we use, C, does not allow array boundaries to be determined interactively.

The first packet on the queue will be processed if and only if the Ethernet MAC is not busy processing other packet. The activity of the Ethernet MAC is indicated by normal value (not DISBALE value) of Event 2, 3 or 4. If the Ethernet MAC is not busy, the first packet in the queue will trigger Ethernet MAC to perform the second event, Channel Seizure Attempt. The Channel Seizure Attempt is scheduled after performing carrier sensing and deference mechanism. Owing to the star configuration, carrier sensing of each station can be done globally since the transmission or jamming signal will arrive at all station at the same time. If the bus is sensed idle than a transmission attempt may be scheduled after the 9.6 μs delay.

### 3.2.6.2 Event 2 : Channel Seizure Attempt

This event models the carrier sensing and deferring mechanism of the Ethernet MAC. The following conditions are responsible to trigger this event :

- Scheduling from packet arrival event as mentioned before.
- Scheduling from the same event if the bus is sensed busy either due to collision event (and subsequent jamming mechanism) or a current packet transmission occupying the bus.
- Scheduling from the third event, collision check event, if the station is involved in a collision (due to collision back off algorithm)
- Scheduling from Event 4 if a transmitting station has outstanding packet to transmit.

As mentioned earlier, this event is useful to speed up simulation execution time. However, there is no statistic collected from this event.

### 3.2.6.3 Event 3 : Collision Check

This event mainly models the collision check mechanism of Ethernet MAC. The collision check is a function that is turned on by a station after sensing (locally) that the bus is free and transmitting its packet. As mentioned earlier, collisions are unavoidable in an Ethernet network and usually happen during the vulnerable period or the collision window, which depends on signal propagation time on the given bus media.

This event also consists of several mechanisms: carrier sensing, jamming, truncated binary exponential backoff calculation and deference. As depicted in Figure 3.5, this event can only be activated by a station having passed channel seizure attempt. Any station entering this event will trigger the main program to calculate the collision window given by Formula 3.1. The program then scans and identifies which stations are attempting in the same collision window. These stations are then considered as collided stations and have to voluntarily delay their transmission based on truncated binary exponential backoff algorithm given by Formula 3.3. The calculation is based on station's backoff counter value for current packet, i.e. number of attempts for current packet. For the same packet, this counter is increased by one every collision. If this value reaches 16, the current packet is discarded. Its ID number is released and all related variables are reset. All stations involved in collision are assumed to transmit a jamming signal of 32 bits. The duration of the collision event is given by Formula 3.2.

Other stations having transmission attempt outside this vulnerable period are considered to perform deference mechanism since their carrier sensing mechanism will abort their transmission attempt. This group of stations will wait until the end of current collision duration. They may schedule their attempt after the 9.6 μs delay following bus idle-sensed indication. The scanning process also records the number of stations attempting transmission within current collision window. If this record value is one, it means the current attempting station has seized or acquired the channel successfully and packet departure event is scheduled.

The following statistics are collected in Event 3:

- number of collision
- maximum attempt of each station
- number of discarded packet due to excessive collisions

Number of total collision is used to calculate collision rate in term of collisions/second. It is also used to calculate packet loss ratio, which is expressed as the ratio of the number of packets discarded divided by the sum of the number of packets, transmitted and the number of packets discarded. Both metrics may be measured for whole network or individually for each station.

### 3.2.6.4 Event 4: Packet Departure

This event is activated by a station that has passed collision check and acquired the channel. The following processes occur in this event:

- resetting the number of collisions (backoff counter) for each transmitting station
- resetting the queue full indicator
- maintaining queue size and packet positions in the queue
- maintaining ID list
- scheduling next transmission attempt if a station's queue is not empty.

The following statistics are collected in this event:

- Packet departure time (individual or aggregate) and its related inter-departure time for every station. This inter-departure time will be tested to find out the sensitivity of the null hypothesis at various network utilisation levels.
- Packet Average transmission delay, calculated as the average elapse time between its first attempt and its ending successful transmission (counting packet duration).
- Network utilisation, calculated as the fraction of time the transmission medium is utilised for a given packet size distribution. It is similar to the total number of bits/second transmitted on the channel for a given packet size distribution. The definition of utilisation will be different in case of model validation. The difference will be mentioned accordingly. For testing the null hypothesis, we consider all bits in standard packet size (between 64 and 1518 bytes) as a useful bit for utilisation calculation.
- Total offered load, computed as the sum of all individual offered loads normalised to 10 MBps or as a percentage of 10 MBps.

- Network Throughput, calculated as packet rate in packets per second. It is calculated for aggregate traffic as well as for each station. For individual station it is a measure of station activity level.
- Collision rate, computed as the ratio of the total number of collisions during simulation to the total simulation time.
- Total simulation time

All of the collected statistics are saved in a suitably formatted file for further analysis. Most of the statistics are used for simulation model validation purposes. For the purpose of testing the null hypothesis, the important statistics are packet inter-arrival time, packet inter-departure time and station activity level (throughput in packets/second).

### 3.2.7 Simulation Unit Time

Most of Ethernet protocol parameter values can be expressed in term of bit times. For example, packet preamble preceding every packet transmission is defined as 64 bit times, interframe gap is defined to be 96 bit times, slot time as a unit of backoff delay is defined as 512 bit times. Therefore, we set our simulation model to be based on 1 bit times unit (= 0.1 μs = 100 nano seconds). During simulation run, all time-related variables are calculated by normalising them to $10^7$ , which corresponds to a single bit time. This value will justify the accuracy of our simulation result.

### 3.2.8 Random Number Generation

Random number generation plays an important role in network simulation. It is used to generate traffic processes to drive simulations in several ways. In our simulation model, we use random number generation to generate the following variables:

- packet's inter-arrival time
- Packet size (from specified distributions)
- Integer number used in Ethernet Backoff algorithm (Formula 3.3)

A good random number generator must pass several statistical tests such as independency test or uniformity test (Jain 1991, pp. 460-473; Ripley 1983, pp. 302-319). Press et al. (1994, pp. 274-286) suggested not to use a system-supplied random number generator, such as **rand()** in C language (It is used in Sadiku Ilyas's models (1995)). This is because the algorithm used in **rand()** produces sequential correlation on successive calls. In our simulation, we use **ran1()** as recommended by Press et al. (1994). It is based on Minimal Standard generator with additional shuffle and safeguards to remove low-order serial correlation. We found **ran1()** is suitable for our simulation amongst other random number generators suggested in Press et al. since it is fast and has long period (up to 100,000,000).

To guarantee independency between stations in our simulation, we assign a single generator for every station. No station generates more than 10,000,000 packets, so we believe **ran1()** is a good choice. For more detail about this random number generator, we refer the reader to Press et al. (1994 Chapter 7 and references herein). It is also worth noting that to obtain specified packet-inter arrival time (e.g., Exponential or Pareto distribution) we use inversion technique (Ripley 1983, Press et al. 1994) since it is simple and fast.

### 3.2.9 Simulation Parameters

Some of simulation parameters dealing with Ethernet protocol parameters have been described above. For convenience, we summarised them again in Table 3.1 with unit time of 1 bit times (= 0.1 μs).

Table 3.1 Ethernet Parameters

| Parameter | Value | Remarks |
|---|---|---|
| Inter-frame gap | 96 bit times | Standard |
| Backoff unit time | 512 bit times | Standard |
| Jamming signal | 32 bit times | Standard |
| Packet preamble | 64 bit times | Standard |
| CRC length | 32 bit times | Standard |
| Collision detect delay | 8 bit times | (Smith & Kain 1991) |
| Signal Propagation Delay | 0.0433 bit times/meter | (Wolfberg 1983, p 81), Note 1 |

Note 1. This value is for coaxial cable bus media without repeater. When varying the bus length which longer than 500 meters, repeater delay should be taken into account.

Other simulation parameters are specified by user and summarised in Table 3.2.

Table 3.2 User Parameters

| Parameter | Value | Remarks |
|---|---|---|
| Number of stations | 1 to 50 | Test Parameter (Note 1) |
| Bus Length | up to 1500 meters | Test Parameter (Note 2) |
| Station Activity level | 1 to 200 packet/seconds | Test Parameter (Note 3) |
| Packet Size | 64 to 1518 bytes | Test Parameter (Note 4) |
| Queue or Buffer Size | 1 to 500 Packets | Test Parameter (Note 5) |
| Simulated runtime | 30 to 120 minutes | (Note 6) |
| Total departed packets | 1 to 3 (million) packets | (Note 6) |
| Start-up transient | 10 second simulated time | (Note 6) |

The following notes for Table 3.2 are closely related to the experimental procedures described in Section 5.2.

Note 1: Number of stations attached to the bus will be varied to account for the effect of background traffic on the null hypothesis. Variation of the station number is useful to investigate the sensitivity of the null hypothesis within certain network utilisation level. For more discussion see Section 2.4.1

Note 2: Bus length will be varied to test the sensitivity of the null hypothesis against network parameter of bus length. For more discussion see Section 2.4.1

Note 3: In this simulation study, 3 stations will be selected to represent 3 stations with different station activity level. They are considered to represent low activity station (with 10 packets/second arrival rate), medium activity station (with 50 packets/second arrival rate) and high activity station (with 100 packets/second arrival rate). The remaining stations are divided into two groups, high activity and low activity. Both groups serve as background traffic generator. The low activity stations will mostly be used to represent receiving stations. This is in line with the nature of asymmetric traffic load on real LAN (Senior, Rehal & Wiseman 1992). This asymmetric traffic loading can be modelled by assigning different arrival rates to each station. In this simulation study, we will investigate the effect of the null hypothesis against different station activity levels. The aim is to find out which station activity level is more sensitive to the null hypothesis.

Note 4: Based on measurement results and for simplicity, packet size distribution will be composed of four packet sizes with different compositions, i.e., two sort packets and two long packets: 64 bytes, 270 bytes, 1080 bytes 1518 bytes. These compositions can be considered as a bimodal length distribution with each mode consists of two kinds of packet sizes. This composition approximates the real Ethernet traffic distribution we observed in our measurement as well as in other measurement studies (Boggs, Mogul & Kent 1995). In our simulation, the packet size will be selected randomly while preserving the specified overall ratio. For more discussion see Section 5.3 and Section 6.3.

Note 5: As discussed in Section 2.4.3, station transmit-buffer size could affect station transmission characteristic. We will investigate the sensitivity of the null hypothesis against this protocol parameter.

Note 6: Either simulated runtime or total number of departed packets on the bus will be used as simulation termination parameter. Given the limited time and computing resources, we need to limit the simulated run time as well as data file size. However, any simulation study should consider the presence of transient state which is normally not of interest (Jain 1991, p 428). Normally observations are made in steady state region. For the purposes of testing the null hypothesis, we do not need this condition since our aim is to compare two related distributions (packet inter-arrival and inter-departure time distribution) regardless of their true distribution. However, to find out the region of applicability of the null hypothesis, we need to perform our observations in steady-state region. For low network utilisation level, simulated runtime will dominate the termination time while the total number of departed packets will dominate at high utilisation levels. See Section 3.4 for more detail about transient removal techniques and simulation stopping criteria in relation to confidence interval of simulation results.

### 3.2.10 Output Statistics

Output statistics and their functions are summarised in Table 3.3.

Table 3.3 Simulation Output Statistics

| Statistics | Function | Type |
|---|---|---|
| Packet inter-arrival time | Test | Individual |
| Packet inter-departure time | Test | Individual and Aggregate |
| Station throughput | Test | Individual |
| Total arriving packet | Evaluation | Individual |
| Total departed packet | Evaluation | Individual and Aggregate |
| Total offered packet | Evaluation | Individual and Aggregate |
| Aggregate departed-packet time series | Self-similarity test | Aggregate |
| Total offered load | Validation | Aggregate |
| Network utilisation | Test and validation | Aggregate |
| Network throughput | Validation | Aggregate |
| Average transmission delay | Validation | Aggregate |
| Packet loss | Evaluation | Individual and Aggregate |
| Collision rate | Evaluation | Aggregate |

Note: The function of *Test* means that the statistic is used in testing the null hypothesis. The function of *Evaluation* means the statistic is used in result discussion or evaluation. The function of *Validation* means the statistic is used in simulation validation (Section 3.3). The function of *Self-similarity* test means the statistic is used in second order statistic tests to test self-similarity in aggregate traffic data.

### 3.2.11 Model Features and Limitations

The developed simulation model implements the following features:

- Packet generator which models packet stream from the Ethernet's client layer. Packet inter-arrival time is generated based on a specified inter-arrival time distribution or source model. The generator model at each station is driven by its own-independent random number generator.

- Each generated packet is placed in FIFO queue and processed in first come first served (FCFS) basis.
- Carrier Sensing, Collision Detection and Jamming mechanism from physical layer.
- Interframe gap timing of 9.6 μs for deference mechanism.
- Truncated Binary Exponential Backoff algorithm with maximum attempt limits of 16 with 512 bit times unit
- Standard packet size ranging from 64 bytes to 1518 bytes with 8 bytes preamble preceding each transmitted packet. Packet size distribution is based on real packet size from measurement.

To reduce simulation complexity while maintaining simulation speed and accuracy, some simplifying assumptions have been incorporated in simulation model. These lead to the following model limitations:

- The model only simulates Ethernet MAC with one way data flow at the transmission site. The receiving processes are assumed to be capable of always receiving the transmitted packets. Boggs, Mogul and Kent (1995) found that the throughput between a pair of communicating stations is limited by the rate at which a single host can send packets. In this study, the receiving process is not an issue. In real Ethernet network, packet receiving processes may affect packet transmission processes through interaction processes occurring in Ethernet controller such as sharing of memory buffer, processor time etc. (Smith & Kain 1991). However, regarding packet transmission processes, this interaction processes should enervate the null hypothesis since it may increase packet delay and affect packet inter-departure time distribution.

- The transmission medium is a noiseless channel. In other words, transmission errors and station malfunction are not considered (Gonsalves 1987, pp. 383-410). The only error is due to the collision event.
- All successfully departed packets are assumed to be received by their destinations without error. In other words, protocol functions like adding headers and trailers by a sending station and checking those fields by a receiving station are not implemented. There is no address destination assigned in each packet. A station may transmit packets to other stations randomly.
- The distance between stations is the same (Star configuration for fairness)
- Independence between packet arrival time and packet size generation.
- The model does not account any software delays in the device driver (Smith & Kain 1991). This delay will have greater effect on short network where end-to-end propagation delay is much lower than packet processing delay. Once again, this should result in better throughput as compared to real system.

### 3.3 Simulation Model Verification and Validation

Being a representation of a system under consideration, a simulation model generally simplifies the system up to certain detailed level to permit valid examination to be drawn about the system. Generally, a more detailed simulation model makes fewer assumptions. However, a more detailed simulation model also means more time to develop, more time to debug and more time to run. Hence, several assumptions are normally unavoidable in model development. These assumptions in turn should be implemented correctly in simulation programs. In case of Ethernet model, the goodness of Ethernet simulation model can then be measured by the closeness of the model output to that of the real Ethernet system.

Validation techniques deal with the correctness of assumptions while verification techniques concern with the correctness of implementation of those assumptions used in model development. Several techniques exist for the purposes of model verification and validation (Jain 1991, pp. 413-422).

#### 3.3.1 Simulation Model Verification

Verification techniques are also known as debugging. They are intended to ensure that the model does what it is supposed to do. By doing so, errors are minimised. We employee the following verification techniques in our simulation programs:

#### 3.3.1.1 Top-Down Modular Design

Our simulation model was designed in several modules. As mentioned before, one event may consists of several sub-events. These sub-events are composed of some procedures and routines that communicate via some interfaces. In addition, we also employee top-down design concept. For example, the packet generator module was composed by two sub-modules, i.e., the inversion function and the random number generator function. By doing so, our program can be easily debugged and maintained.

#### 3.3.1.2 Antibugging

To verify our simulation program, we have included several check points to detect the presence of the bugs. For example, we check the aggregate inter-departure time between packets to find out the present of overlapping packets. If the overlapping packets are detected, it means there is something wrong with the algorithm and the program should be modified accordingly. Other check points include verifying packet size, verifying number of valid events and verifying number of valid ID list.

#### 3.3.1.3 Event Tracing

We employed manual tracing of events during simulation. It is possible through facility provided by The Integrated Development Environments (IDE) of Borland C++ 3.1 compiler (McCord 1992, pp. 12-21).

### 3.3.1.4 Consistency Test and Seed Independence

We test our simulation result for consistency of output and independency of seed. The simulation was executed twice with 10 stations and 20 stations respectively for about 33 minutes simulated runtime for each execution (limited to 1 million departed packets). Packet inter-arrival time is set to be exponential with the same rate and buffer for each station. The packet size distribution is bimodal with 75 % of 64 bytes and 25 % of 1518 bytes. In the first run, each of 10 station was assigned 50 packets/second arrival rate while in the second run, each of 20 station was assigned 25 packets/second arrival rate. The resulted network utilisation (in MBps) for the same packet size distribution for both simulations is very close. The difference was not significant for several executions with different seed. Table 3.4 summarises the test results.

### 3.3.2 Simulation Model Validation

The rationale behind validation techniques is that if the assumptions are correct and correctly implemented in simulation programs then the model would give results that are close to those of real systems. Practically, this will involve comparison studies followed by justifying arguments about the differences between the model output and the discussant since errors should normally be present. Unlike verification techniques, validation techniques are model-dependent. Therefore, to validate our simulation model we surveyed several published works on Ethernet measurement as well as analytical studies.

Table 3.4 Consistency Test Statistic

| Execution | Number of Hosts | Arrival Rate | Utilisation (MBps) |
|---|---|---|---|
| 1 | 10 | 50 | 1.7405 |
| 2 | 10 | 50 | 1.7412 |
| 3 | 10 | 50 | 1.7363 |
| 4 | 10 | 50 | 1.7427 |
| 5 | 10 | 50 | 1.7413 |
| 6 | 20 | 25 | 1.7419 |
| 7 | 20 | 25 | 1.7377 |
| 8 | 20 | 25 | 1.7430 |
| 9 | 20 | 25 | 1.7421 |
| 10 | 20 | 25 | 1.7436 |

Unfortunately, there is no analytic model of Ethernet network currently available (Gonsalves 1987, pp. 383-410), but there exist several published analytic models of CSMA/CD networks based on star configuration. These models differ from Ethernet mainly in retransmission strategy used to handle collision event. Analytic models of CSMA/CD simply assumed exponential retransmission delay time upon a collision. This assumption causes significant performance degradation at high offered load since calculation of simple exponential retransmission delay does not involve any knowledge of the current load condition of the bus. In other words, CSMA/CD analytical models use uncontrolled retransmission strategy. Typical throughput versus offered load curve will have one maximum mode (Tobagi & Hunt 1987, pp. 318-339; Sohraby, Molle & Venetsanopoulos 1987; Takagi & Kleinrock 1987). After this point, throughput will drop rapidly (unstable condition) as offered load increase. On the other hand, Ethernet's throughput versus offered load curve will stabilise after reaching its maximum value (Metcalfe & Boggs 1976; Gonsalves 1987, pp. 383-410). At extreme overload conditions, Ethernet utilisation remains high and stable; it has no signs of suddenly decreasing. This is because a controlled retransmission strategy used by Ethernet is designed to dynamically estimate network load (by monitoring the number of collisions experienced by any one packet) and adjusts the retransmission delay accordingly. However, despite the difference between their retransmission strategy, CSMA/CD analytical models are still useful for the purpose of model validation.

We will therefore use CSMA/CD analytical models as well as measurement results to validate our model. The following statistics and published work are used to validate our simulation programs. They can be classified into two groups: fixed station number with variable station's offered load and fixed station's offered load with variable number of stations.

1a. Utilisation versus Offered Load for fixed packet size(Gonsalves 1987, pp. 383-410).

1b. Utilisation versus Offered Load for fixed packet size (Sohraby, Molle & Venetsanopoulos 1987, pp. 240-243; Takagi & Kleinrock 1987, pp. 243-245).

2a. Utilisation (MBps) versus Number of Host for fixed packet size (Boggs, Mogul & Kent 1995).

2b. Packet Rate (Packets/second) versus Number of Host for fixed packet size (Boggs, Mogul & Kent 1995).

2c. Utilisation (MBps) versus Number of Host for variable packet size (Boggs, Mogul & Kent 1995).

2d. Average Transmission Delay (ms) versus Number of Host for fixed packet size (Boggs, Mogul & Kent 1995).

For more detail about the discussant studies used to validate our simulation model, we refer the reader to the references. The following validation results are obtained from 10 execution of each simulation (with different seed) for 200000 departed packets from each simulation. The statistics are collected after removing the start-up transient (see Section 3.4). The validation metrics is then calculated for the 95 % confidence interval with 9 degrees of freedom (see Sadiku & Ilyas 1995, pp. 72-77).

#### 3.3.2.1 Validation Results

1. Validation to measurement and analytical studies with variable station's offered load

Gonsalves (1987, pp. 383-410) conducted a set of experiments to study the performance of Experimental (3 MBps) and Standard (10 MBps) Ethernet. In particular, he did a comparison study between his experimental results and published CSMA/CD analytical studies to study the possible maximum throughput. Gonsalves's experimental parameters that we simulated for the purpose of validation are: homogenous 32 stations, 11.75 μs end-to-end delay, 1 packet buffer for each station and fixed packet size of 64 bytes, 512 bytes and 1500 bytes. In

Gonsalves's measurement, the 4 bytes frame check sequence is excluded from throughput calculation. We adjust our simulation program accordingly.

To obtain numerical data from Gonsalves's measurement, we have utilised a graph in his paper. The graph is magnified and numerical data is extracted from enlarged graph. The same technique is applied to data of the second measurement study (Boggs, Mogul & Kent 1995) used to validate our model since the numerical data is not available.

Since our simulation model is unslotted model, we prefer to use unslotted CSMA/CD models proposed by Sohraby, Molle and Venetsanopoulos (1987, pp. 240-243) that are equivalent to the model of Takagi and Kleinrock (1987, pp. 243-245). For analytical models, we set the parameters to be the same with our simulation model which result in value of **a** (i.e., the normalised end-to-end propagation delay to the packet duration) of 0.2295, 0.02869 and 0.0098. This values correspond to the end-to-end propagation delay of 11.75 μs and the packet size of 64 bytes, 512 bytes and 1500 bytes respectively. We set the **b** parameter of their models to be 0.

For the sake of simplicity, we use the computationally simpler infinite population formula since for large populations, the analysis result for finite and infinite populations will converge (Gonsalves 1987, pp. 383-410). Figure 3.8 summarises this validation result. Note that the errors calculated with the 95 % confidence interval and 9 degrees of freedom are very small compared to the average values. For example, the errors of the simulation results of 512 bytes packet size are between 9.7E-05 to 8.97E-04 (for throughput) which are 3 to 4 orders of magnitude smaller than their average values. For this reason, all graphical presentations will be presented based on their average values only without their confidence limits.

Despite the differences, our simulation results show a good stability as expected by previous measurement studies. The maximum throughput predicted by CSMA/CD model also indicated a good agreement with our simulation. Compared to Gonsalves' measurement results, our simulation resulted in higher utilisation while Gonsalves's measurement show lower utilisation especially for shorter packets. One possible cause of this discrepancy is the performance of the Ethernet driver used in Gonsalves's experiments. It is shown by Minnich and Cotton (1983), in their evaluation study, that the Ethernet Controller transmission throughput from two different vendors may differ significantly depending on their design approach. For short packet, one controller may transmit packets with rate up to 1750 packets/second while other is limited to 500 packets/second. In fact, another measurement study (Boggs, Mogul & Kent 1995), described shortly, revealed that even with short packet size (64 bytes) in about the same network condition, the throughput of 0.8 can be achieved.

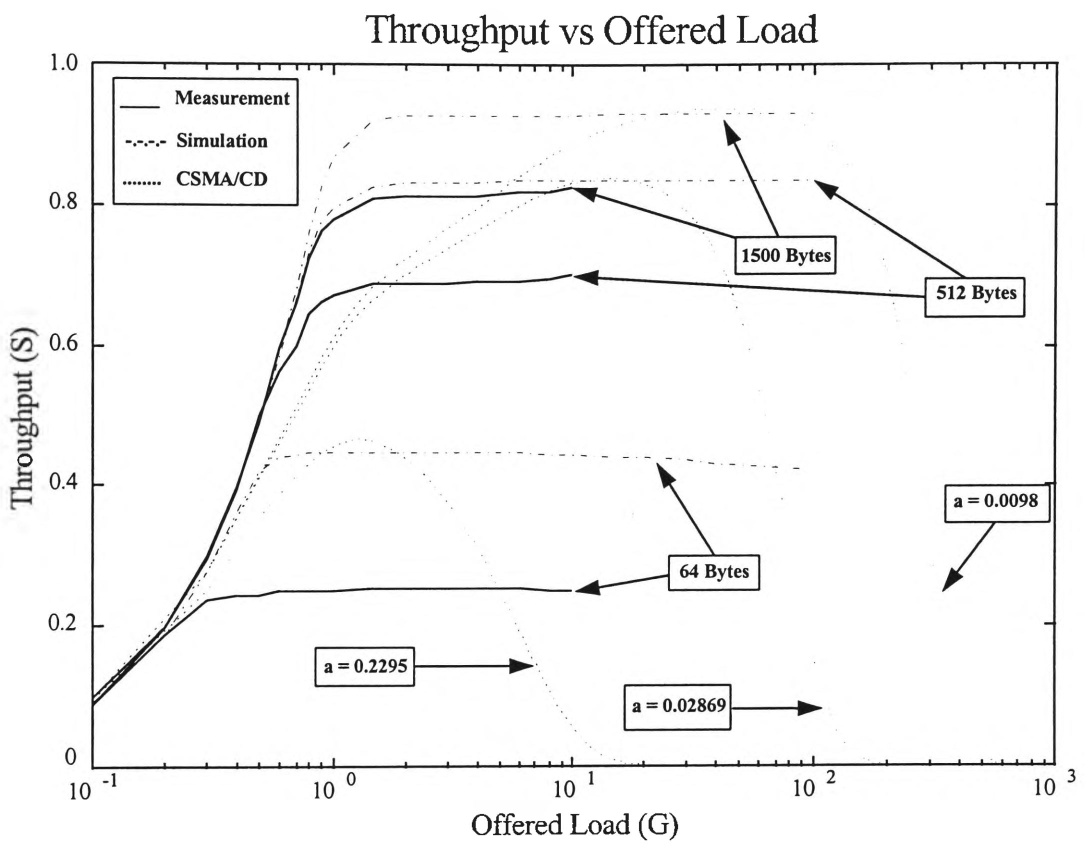


Figure 3.8 Utilisation versus Offered Load for Fixed Packet Size

Our simulation result presented in Figure 3.8 was obtained using exponential inter-arrival time. However, we also conducted the simulation for the same parameters using Pareto inter-arrival time with 1.5 shape parameter (see Chapter 5 for more detail). The results (not shown here) are the same with those of exponential inter-arrival time. This required the simulation to be 3 times longer due to higher variability of Pareto inter-arrival time.

2. Validation to measurement with fixed station's offered load

The second measurement study we used to validate our model is the study of Boggs, Mogul and Kent (1995). Basically, they investigated the possible maximum capacity of Ethernet under varying combinations of packet lengths, network lengths and number of hosts. In doing so, they set the hosts to continuously generate a total offered load that exceeds the capacity of the network. In our simulation, this requires the traffic intensity (Formula 3.4) to be always slightly greater than 1 while varying other parameters.

We selected several measurement results to validate our simulation model. Since the available measurement results were obtained from linear bus networks, we selected the measurement results from network configurations that have about the same condition with our simulation configurations. They can be grouped into two statistics: fixed and variable packet size statistics. The fixed packet size statistics include: utilisation (total bit rate) versus number of hosts for short network, total packet rate versus number of hosts for short network (6 meters) and long network (910 meters), total packet rate for long network and average transmission delay for long network. For variable packet size the selected statistic is utilisation versus number of hosts. Two bimodal packet size distributions are used for this validation purpose.

We only validated our simulation program for 3 packet sizes and for 4 kind of station numbers: 10. 15, 20 and 23 (for short network) or 24 (for long network).

There is different definition of utilisation in the second measurement study. To calculate utilisation or total bit rate, each packet size is charged with 20 bytes extra to account for the 9.6 μs interframe gap (12 byte times) and the 8 byte times of preamble (The 4 byte frame check sequence is part of packet size in our simulation). It is intended to yield maximum bit rate of 10 MBps when the network is carrying back-to-back packets without collisions.

Since our model represents star configuration, we may expect the measurement results for short network to have about the same result with our simulation. We set our simulation parameters as follows: 100 packet buffers at each station, 0.03 μs end-to-end delay (correspond to 6 meters bus length) and exponential inter-arrival time. We also run the same simulation for Pareto inter-arrival time for cross-checking purposes. The results for both arrival model did not differ significantly. Only the result of exponential arrival model is presented. Figure 3.9 depicts the validation result for short network.

Our simulation results indicated that in short network the measurement utilisation are slight lower than simulation ones, but the trend is in good agreement. One possible cause of the discrepancy is the present of inherent delays (e.g. software delay, interrupt delay or data copy delay) in real Ethernet driver which have been ignored in our simulation program. These delays determine the upper bound of the Ethernet controller transmission throughput (Minnich & Cotton 1983; Smith & Kain 1991).

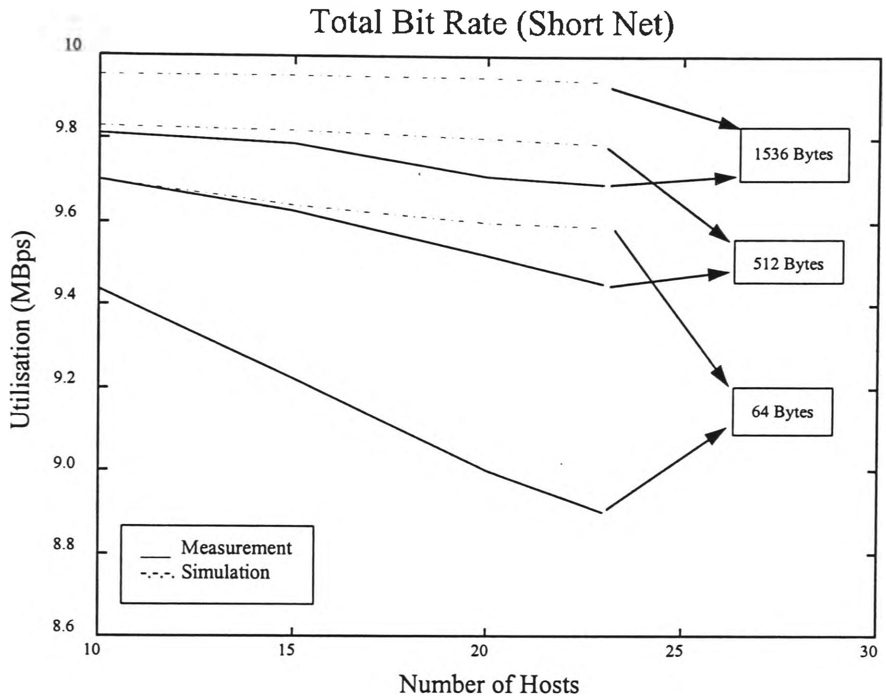


Figure 3.9 Utilisation versus Number of Hosts for Fixed Packet Size (Short Net)

On the other hand, when we compare the packet rate using the long network (910 meter) which equals to 10 μs end-to-end delay (Boggs, Mogul & Kent 1995), our simulation results indicated the worst-case performance since in a star configuration the distance between stations is the same. The results of this validation are presented in Figure 3.10. As depicted in Figure 3.10, the measurement results are slight higher than the simulation results, but for longer packets our simulation results match measurement results very well. For short packet size (64 bytes), when the packet transmission time is only an order of magnitude larger than the collision window, the linear bus configuration (as used in the measurement study) will produce higher packet rate because collisions get resolved quicker.

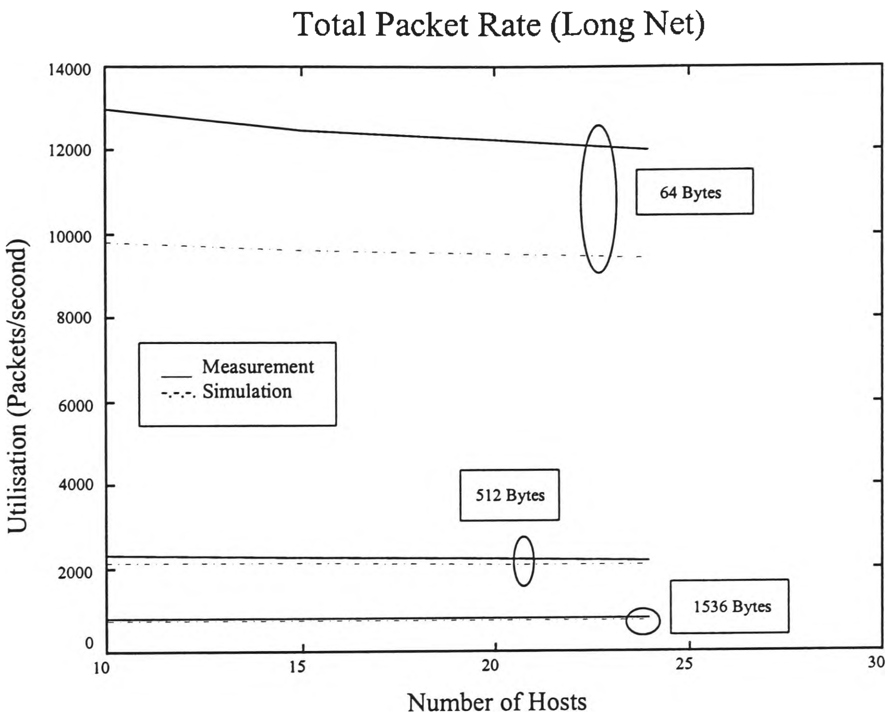


Figure 3.10 Packet Rate versus Number of Hosts for Fixed Packet Size

We also validated our model with variable packet size as presented in Figure 3.11. We use two distributions of packet size that labelled as 7/1 and 6/2. The distribution of 7/1 means the ratio of short packet (64 bytes) to long packet (1536 bytes) is 7/8 to 1/8. The distribution of 6/2 is defined accordingly. During the simulation run, the packet size is selected randomly while preserving the overall ratio. As in the previous results, our simulation reveals a good trend agreement. The lower utilisation of our result is the worst-case performance.

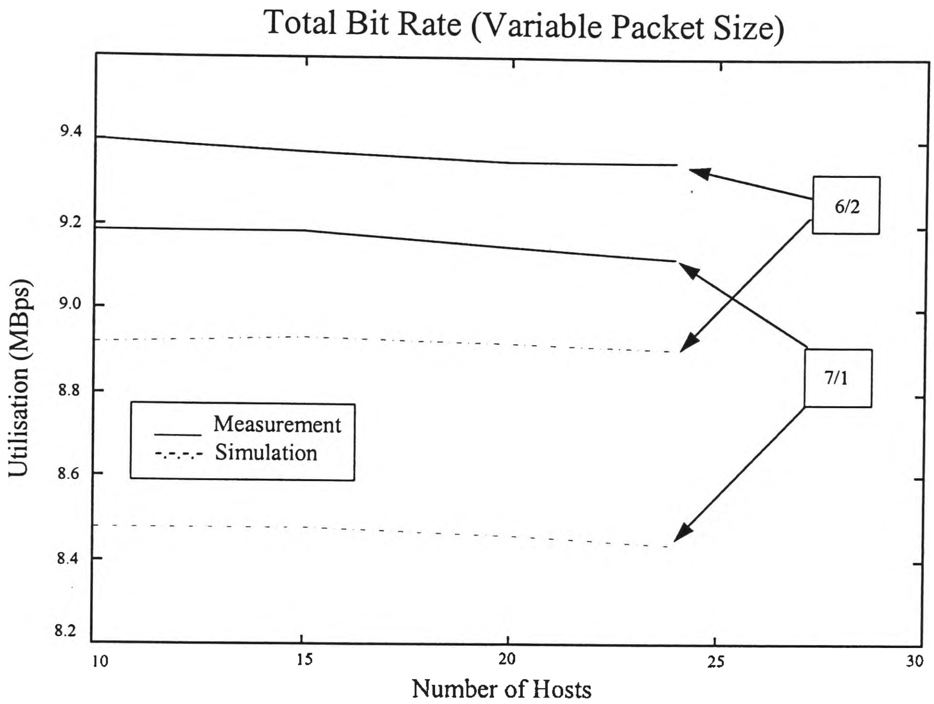


Figure 3.11 Utilisation versus Number of Hosts for Variable Packet Size

To complete the validation task, we compare the average transmission delay of the measurement result to our simulation result. As defined earlier, the packet transmission delay is defined as the elapsed time between its first attempt to be transmitted to the end of its successful transmission thus counting all bytes in packet (from 64 bytes to 1518 bytes, excluded packet preamble). This is in accordance with the definition used in (Boggs, Mogul and Kent 1995). The validation result is presented in Figure 3.12.

For long packet size, the delay of measurement results is differ significantly from simulation result. We believe it is due to the discrepancy in definition of packet first attempt. In our simulation model, packet first attempt is defined as the starting time of Event 2 thus exclude packet processing time and packet queuing delay. Other factor that may responsible for this discrepancy is delay due to Ethernet controller interrupt used to measure the end of packet transmission. In their study, Smith and Kain (1991) found that the interrupt form the LANCE to the CPU to indicate the end of packet transmission might be delayed whenever higher priority interrupts are already in service. However, the trend of our simulation result showed a good agreement with measurement one.

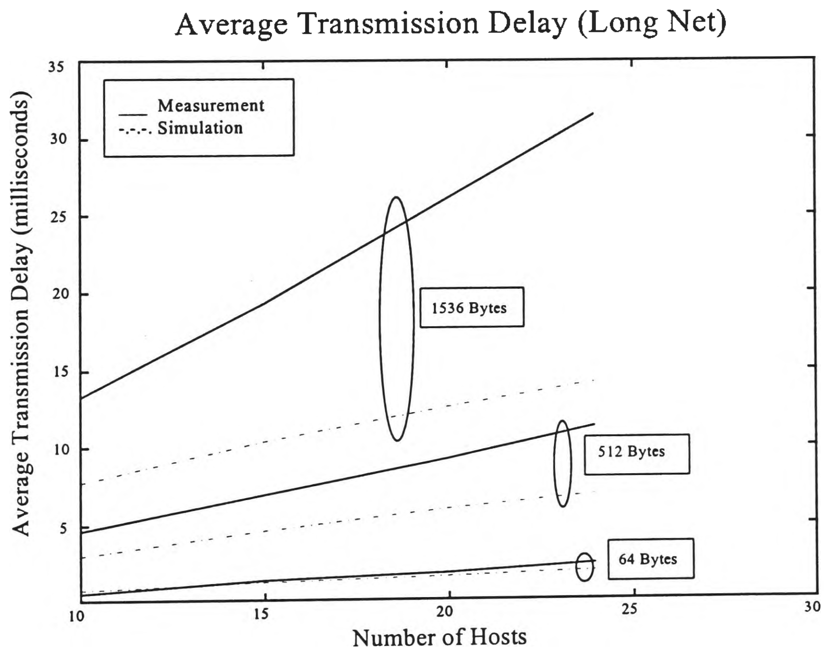


Figure 3.12 Average Delay versus Number of Hosts for Fixed Packet Size

### 3.4 Transient Removal and Stopping Criteria

Under given simulation configurations and parameters, it is interesting to investigate the applicability range of the null hypothesis. For this purpose we need to run the simulation in its stable state before observations can be made. In this case, transient data is not of interest. To minimise the influence of transient data, we utilise long runs technique combined with initial data deletion (Jain 1991, pp. 423-430). We use long runs technique because one of the source generator models that will be used (Pareto distribution, see Chapter 5) needs long simulation times before the statistics approach equilibrium (Gordon 1996, pp. 28-39). In our simulation, the transient data is indicated by the domination of the first station having the first chance to acquire the channel. After several investigation, we found that after 2 seconds of simulated time, this transient behaviour is stabilised.

The length of simulation time needs to be selected properly to avoid wasted computing resources while maintaining the accuracy of the results This means simulation time must be neither too short nor too long. Simulation stopping criteria thus becomes another important subject in simulation studies. In line with transient removal technique we described above, we employ independent replications technique for simulation stopping criteria (Jain 1991, pp. 430-432). Each simulation will be executed 5 times with different seed to verify the bias in simulation results. Each execution produces at least 1 million packets. With this figures the accuracy of our simulation result is justified. As a comparison, Gonsalves And Tobagi (1988) used the method of sub-runs or batch means (Jain 1991, pp. 432-433) with 10 consecutive sub-runs (each of 1 second duration) to obtain the 95 % confidence intervals within 1 % of their aggregate statistics. For low utilisation levels our test showed that this equals no more than 10,000 departed packets in aggregate data.

## 3.5 Summary

This chapter has presented the development, verification and validation of the Ethernet simulation model. The Discrete Event Simulation technique has been used to develop the simulation model. The model was designed based on the proposed CSMA/CD model with several improvements to account for the standard Ethernet protocol parameters. Several underlying assumptions based on the previous studies were used to facilitate model development.

Several important aspects of simulation studies, such as random number generation, model verification, model validation, transient removal technique and stopping criteria have also been described and incorporated in the Ethernet simulation model. The simulation validation results have shown that the developed model is in good agreement with real Ethernet network.

# 4. Statistical Methods

## 4.1 Introduction

The previous chapter described the simulation program used to generate statistical data of packet inter-arrival and inter-departure time distributions. The null hypothesis that these 2 distributions are the same will be tested in the next chapter. The statistical methods required to test the null hypothesis are described in this chapter. This chapter thus completes the requirements for testing the null hypothesis.

Section 4.2 presents graphical methods used to analyse simulation data. These methods serve as complementary methods for the purposes of testing the null hypothesis since they are not formal methods. The formal methods for testing the null hypothesis are described in Section 4.3. They are actually tests based on empirical cumulative distribution function (EDF).

The chapter is concluded by presenting second order statistic tests in Section 4.4. These statistics are used to test the presence of self-similarity in aggregate traffic data. The aim is to give statistical reference when comparing two simulation results of Poisson traffic source model and Pareto traffic source model. This short section is intended as an introduction only since the self-similarity is not the main issue of this thesis.

## 4.2 Graphical Methods

To test the null hypothesis, graphical methods are used in conjunction with the formal numerical methods (D'Agustino & Stephens 1986, p 7). The use of graphical methods alone in data analysis should be avoided since it may lead to unjustified conclusions. Therefore, formal numerical techniques are often essential to avoid this. However, graphs can sometimes reveal hidden characteristics of data that were not predicted prior to the analysis. For these reasons, we employee two graphical methods known as empirical cumulative distribution functions (ECDF or EDF for short) and probability plotting technique based on EDF known as Quantile plots (Q-Q plots). Basically, we use these graphical techniques to compare the distribution of packet inter-arrival time (IAT) distribution and packet inter-departure time (IDT) distribution. For more information on EDF and probability plotting techniques, we refer the reader to the following references: (D'Agustino & Stephens 1986, pp. 7-62; Fisher 1983; Gerson 1975; Gnanadesikan 1977, pp. 197-200; Kotz & Johnson 1985; Wilk & Gnanadesikan 1968).

### 4.2.1 Empirical Cumulative Distribution Function (EDF)

Empirical cumulative distribution function (EDF) is a primitive technique that is usually used to compare two distributions (Wilk & Gnanadesikan 1968). It is a step function, calculated from the sample that estimates the distribution function of the sample. Given the random sample of $X_1, X_2, \ldots, X_n$ drawn from a distribution with cumulative distribution function (CDF) $\mathbf{F}$, then the EDF is defined as (D'Agustino & Stephens 1986, p 8):

$$F_n(x) = \#(X_j \leq x)/n, \quad \infty < x < \infty \qquad \textbf{(Fml. 4.1)}$$

where $\#(X_j \leq x)$ is the number of $X_j$ less than or equal to $x$. The use of ≤ sign has lead to the other name of EDF i.e., less than cumulative distribution function (LTCDF). Accordingly, the distribution of $1 - F_n(x)$ is known as Complementary EDF (CEDF) or more than cumulative distribution function (MTCDF). The EDF plot is useful to compare the characteristic on lower tail of distributions while CEDF is used to compare upper tail of two distributions.

The EDF is a non-parametric technique since its usage does not involve any assumptions dealing with the underlying parametric distributions. This technique does not require the grouping task that is an imperative in histogram-based techniques. But, for a large sample size, the difficulty arises in ordering the data, either in ascending or descending order before the data is plotted. However, the availability of good sorting algorithm (Press et al. 1992, pp. 329-346) and given the speed of current computer technology this problem is minimised.

To obtain more information from EDF plotting, we use a logarithmic (log-log) scale instead of a linear scale. The difference in using two different scales on the same distribution data is depicted in Figure 4.1 and Figure 4.2. In Figure 4.1 we use linear scale while in Figure 4.2 we use logarithmic scale. From these 2 figures, it is apparent that EDF plotting with logarithmic scale is more powerful than the one with linear scale. As we will show in the next chapter, it is this kind of discrepancy that has lead to the rejection of the null hypothesis.

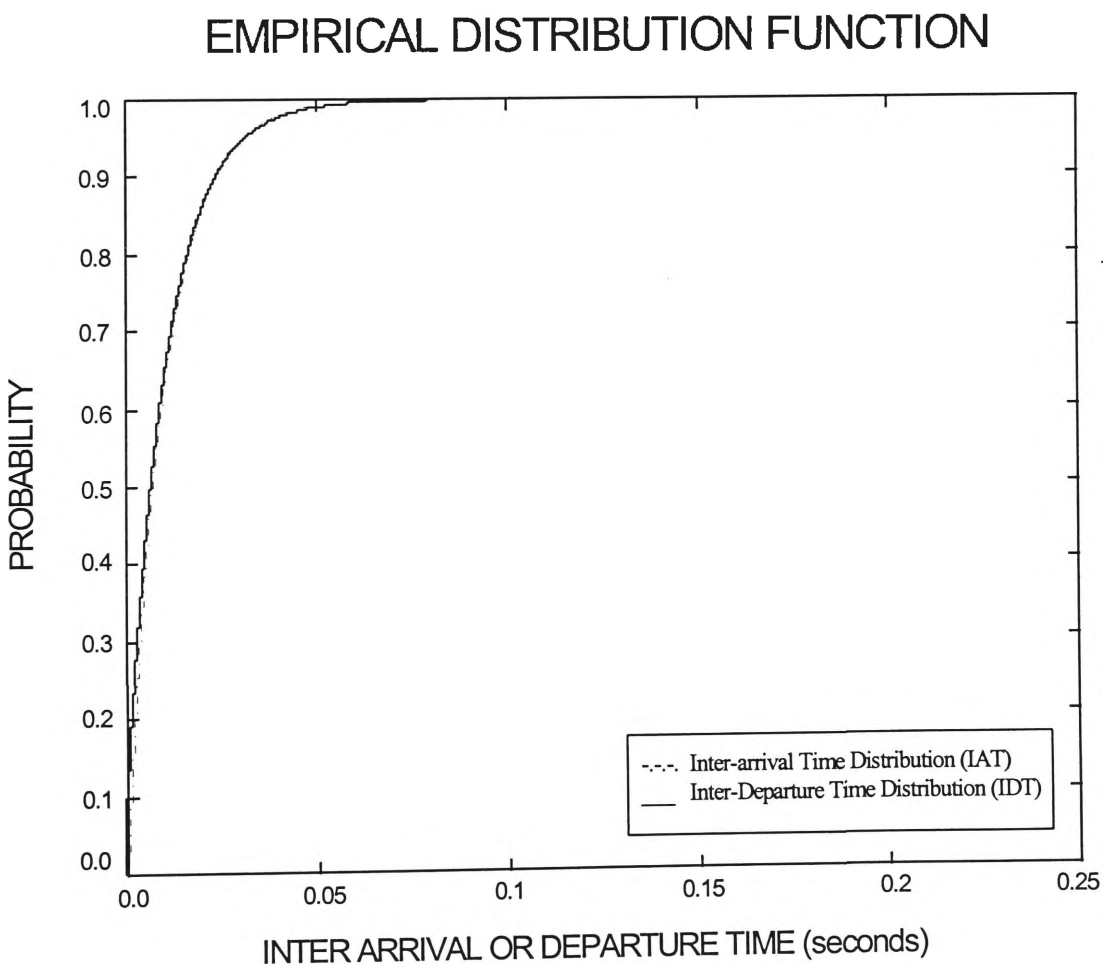


Figure 4.1 EDF Plot with Linear Scale

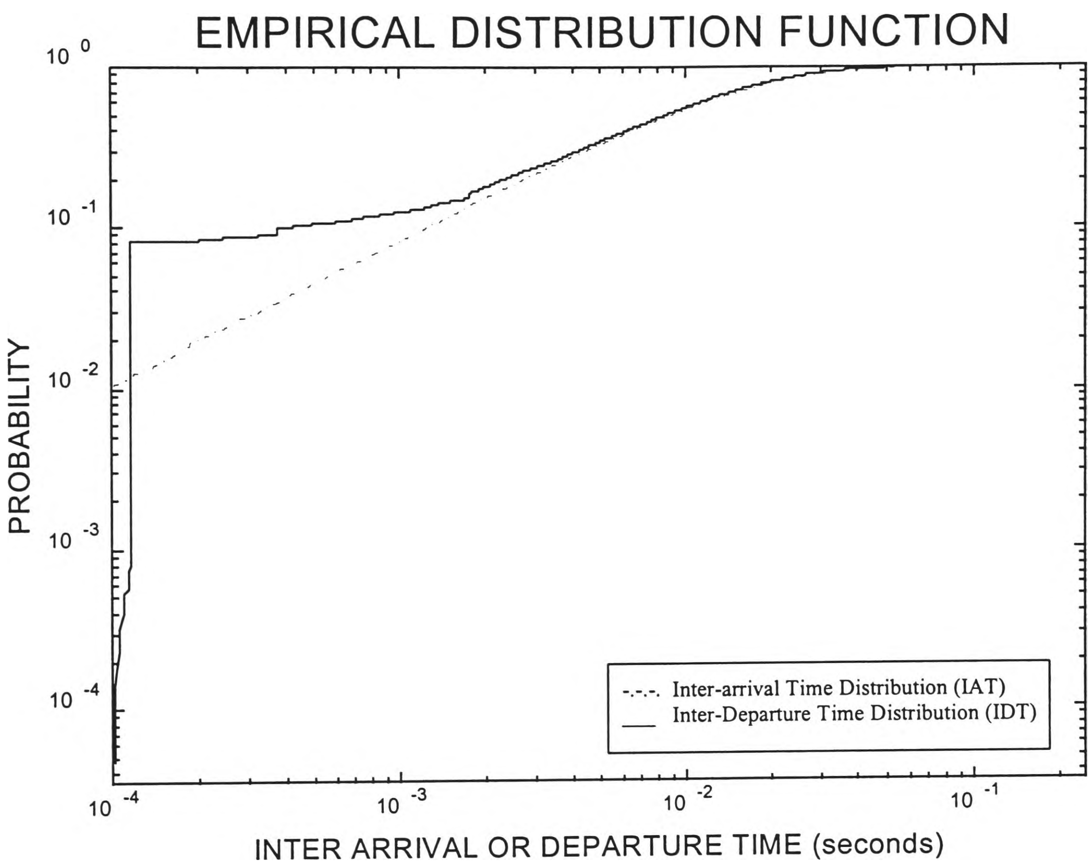


Figure 4.2 EDF Plot with Logarithmic Scale

### 4.2.2 Q-Q Plot

The comparison technique of two distributions based on their EDF like those presented in Figures 4.1 and 4.2 can also be performed by plotting their ordered quantiles against each other. The resulting plot is known as Quantile-Quantile (Q-Q) plot. If two distributions are the same then the Q-Q plot will be a $45^0$ line through the origin. The straight line of Q-Q plot can be used to calculate the location and scale parameters of one distribution relative to the other (Gerson 1975). If one distribution is of the standard form, the gradient of the line will give the value of standard deviation while the mean is given by the point at which the line intercepts the y-axis. However, as mentioned earlier, the Q-Q plot is more useful for the presentation purpose since human eye can readily perceive a straight line and identify the slope and intercept. But, these plots can not give the significance level at which we have to accept or reject the null hypothesis. Therefore, to test the null hypothesis we will utilise the formal numerical methods as described below.

### 4.3 Formal Method: Kolmogorov-Smirnov Test with V Statistic

In this study the problem of testing the null hypothesis can be considered as the question of whether the two data sets (i.e., IAT and IDT) are drawn from the same distribution functions, or from different distribution functions. Statistically we need to find out, to a certain required level of significance, whether we can disprove the null hypothesis that the two data sets are from the same distributions. Disproving the null hypothesis at a certain level of significance means the null hypothesis is rejected at that significance level. On the other hand, failing to disprove the null hypothesis, i.e., the null hypothesis is accepted, shows that the two data sets are from the same distribution functions.

Since IAT and IDT data are continuous variables, the most generally accepted test is the Kolmogorov-Smirnov (K-S) test and its derivatives (Press et al. 1992, p. 620). This test is based on EDF statistics which are defined as measures of the discrepancy between the EDF of the two distributions under consideration

(D'Agustino & Stephens 1986, p 97). These statistics are based on the vertical differences between two EDFs. The power of EDF-based statistics are usually higher than the Chi-squared statistics (D'Agustino & Stephens 1986, p 110; Porter III, Coleman & Moore 1992).

There are two classes of EDF-based statistics i.e., *supremum* and *quadratic* (D'Agustino & Stephens 1986, p 100). Belonging to the first class is the most well known D statistic introduced by Kolmogorov and the V statistic proposed by Kuiper. The quadratic class is known as Cramér-von Mises statistic family which consists of Cramér-von Mises $W^2$ statistic, Anderson-Darling $A^2$ statistic and Watson $U^2$ statistic.

Our preliminary study shows that most of the discrepancies between IAT and IDT distributions are on the tail, either lower or upper. The recommended statistics to detect any departure in the tails is Anderson-Darling $A^2$ statistics (D'Agustino & Stephens 1986, p 110). However, in case of two samples, as in this thesis, there is no simple formula that relate the value of $A^2$ statistics to its significance level (Press et al. 1992, pp. 626-627). Therefore, to test the null hypothesis we use the V statistic as suggested by Press et al. (1992, pp. 626-627). Compared to D statistic, the V statistic is more powerful to test the discrepancy on the tail of distributions (Press et al. 1992, p 627). Furthermore, there exists a simple formula for the asymptotic distribution of the statistic V that relates the value of V statistic to its significance level.

Kuiper's V statistic is defined as (D'Agustino & Stephens 1986, p 100):

$$V = D^{+} + D^{-} \qquad \textbf{(Fml. 4.2)}$$

where $D^{+} = sup_x|F_1(x) - F_2(x)|$ and $D^{-} = sup_x|F_2(x) - F_1(x)|$. In our case, $F_1(x)$ is the EDF of inter-arrival time and $F_2(x)$ is the EDF of inter-departure time. For more information on K-S test used in this thesis we refer the reader to the books of Press et al. (1992 pp. 623-627) and (D'Agostino & Stephens 1986, pp. 100-101). As reference, for all the K-S tests of the null hypothesis we also calculate the D statistic and its significance level.

## 4.4 Second Order Statistics for Testing Self-Similarity in Aggregate Traffic

### 4.4.1 Self-Similar Stochastic Processes

The self-similar nature of Ethernet traffic data as coined initially by Leland et al. (1993) has attracted growing interest in traffic modelling research. This nature is closely related to the statistical property known as *long-range dependence (LRD),* as opposed to *short-range dependence (SRD)* or independence property (Hampel 1987). The proposed traffic models of self-similar processes are capable of capturing burstiness observed in real traffic data (Leland et al. 1993). The burstiness of aggregate traffic data can be described by a single parameter H known as *Hurst* *parameter.* This nature of self-similar traffic models may facilitate a simple, accurate and realistic description of traffic scenarios in B-ISDN deployment.

A continuous time stochastic process is called self-similar with parameter $H$, if for any $k$ and any $k$ times points, $t_1, t_2, \ldots, t_k$ the joint distribution of $X(t_1), X(t_2), \ldots, X(t_k)$ is the same as that of $a^{-H}$ times the joint distribution of $X(at_1), X(at_2), \ldots, X(at_k)$ for any $a > 0$ (Hampel 1987). Intuitively, the process looks the same in every time scale. In practice we usually use the discretised version of this stochastic process as observed at integer time points.

**Definitions**

Let $X = \{X_k : k \geq 1\}$ be a covariance stationary stochastic process in discrete time domain with mean $\mu = E[X_k]$, variance $\sigma^2 = E[X_k - \mu]^2$ and an autocorrelation function $r(k) = E[(X_n - \mu)(X_{n+k} - \mu)]/\sigma^2$ depending on $k$ only. In this case $X_k$ represents the number of packets observed in the $k$-th time interval of size $m$ seconds. The process $X$ is said to be a LRD process (Leland et al. 1994) if its autocorrelation function $r(k)$ has the form of:

$$r(k) \sim k^{-\beta} L(k) \qquad \text{as } k \rightarrow \infty \qquad \textbf{(Fml. 4.3)}$$

with $\beta = 2 - 2H$ and $0.5 < H < 1.0$ or $0 < \beta < 1$. $L(k)$ is a slowly-varying function, i.e., $lim_{t \rightarrow \infty} L(tx)/L(x) = 1$ for all $x > 0$. Constants and logarithms are examples of slowly-varying functions (Paxson 1995).

From the original process $X$ we can define an aggregate process

$$X^{(m)} = \{X_k^{(m)} : k \geq 1\}$$

with autocorrelation function $r^{(m)}(k)$ where

$$X_k^{(m)} = (X_{(k-1)m+1} + X_{(k-1)m+2} + \ldots + X_{km})/m. \qquad \textbf{(Fml. 4.4)}$$

If $r^{(m)}(k) = r(k)$ for any $m = 1,2, \quad ,(k=1,2)$ then the process $X$ is called *(exactly) second-order self-similar* Hurst Parameter (Georganas 1994)

$$H = 1 - (\beta/2) \qquad \textbf{(Fml 4.5)}$$

The process $X$ is called *asymptotically second order self-similar* with parameter $H = 1 - (\beta/2)$, if $r^{(m)}(k) \to r(k)$, as $m \to \infty$. This relation depicts a non-summable autocorrelation function, i.e. $\sum_k r(k) = \infty$. On the other hand, the process $X$ is called a SRD process if its aggregated processes $X^{(m)}$ tends to second-order white noise, i.e., for all $k \geq 1$, $r^{(m)}(k) \to 0$, as $m \to \infty$. Accordingly, it implies a summable autocorrelation function $\sum_k r(k) < \infty$. Typical Markovian traffic models considered in literature have geometrically decreasing autocorrelation functions and thus are SRD processes (Leland et al. 1993).

In practice, the present of self-similarity ('fractal') nature in aggregate traffic data can be detected by measuring its Hurst Parameter. The LRD processes will have H value in the range of 0.5 < H < 1.0 while H value of SRD processes is 0.5 (Leland et al. 1994). Several statistical methods for estimating the self-similarity parameter H or the intensity of LRD in a time series are available. In this thesis we use two simple graphical methods known as Index of Dispersion for Count (IDC) or Fano factor (Johnson & Kumar 1990) and Variance Time Plot (VTP). For more information about other statistical methods for estimating the self-similarity parameter we refer the reader to an empirical study of Taqqu, Teverovsky and Willinger (1995).

### 4.4.2 Index of Dispersion for Count (IDC)

The index of dispersion for counts (IDC) is initially used as a measure of burstiness. If $X$ represents the total number of packet seen every $m$ unit time (e.g., every 10 ms) then the IDC is defined as (Leland et al. 1994):

$$IDC(L) = \frac{\operatorname{var}\left(\sum_{j=1}^{j=L} Xj\right)}{E\left[\sum_{j=1}^{j=L} Xj\right]} \sim cL^{(2H-1)} \qquad \textbf{(Fml. 4.6)}$$

where c is a finite positive constant that does not depend on **L**. By plotting $log(IDC(L))$ against $log(L)$ we may obtain an asymptotic straight line with slope $2H - 1$. In practice, the slope is estimated by fitting a line to the points obtained from the plot. If the gradient is $g$ then the estimated $H$ parameter is

$$H = \frac{(g+1)}{2} \qquad \textbf{(Fml. 4.7)}$$

Figure 3.2.4 and 3.2.5 in (Leland & Wilson 1991) give a good presentation of the difference between LRD and SRD processes in terms of their IDC curves. The IDC plotting of LRD process is increasing monotonically over a wide range of time scale. In contrast, the IDC value of pure Poisson processes (i.e., SRD processes) is a constant of 1 while Poisson batch processes have an IDC that rapidly converge to small fixed values.

Willinger et al. (1995) showed that to generate self-similar traffic with ON-OFF traffic model, the main ingredient that is needed to obtain $H > 0.5$ is to use the intensity *of Noah Effect* $\alpha$ in the range of $1 < \alpha < 2$. In this case, $\alpha$ can be considered as the shape parameter of a Pareto distribution since Willinger et al (1995) mentioned that one of the ON or OFF period distribution should have a hyperbolic tail or Pareto-like distribution in order to generate self-similar traffic with $H > 0.5$. Willinger et al (1995) gave a simple formula that that relate $\alpha$ and $H$ as:

$$H = \frac{3 - \alpha}{2} \qquad \textbf{(Fml. 4.8)}$$

Since one of our test traffic models is the Pareto model, we are interested in verifying the resultant self-similar nature in our simulation traffic data.

### 4.4.3 Variance Time Plot (VTP)

The variance time plot (VTP) is another graphical method for distinguishing between SRD and LRD. It is obtained by plotting $log(var[X^{(m)}])$ against $log(m)$ (i.e., time interval) where $X^{(m)}$ is the aggregated process defined in Formula 4.4. Then for SRD traffic process we have (Erramilli, Narayan & Willinger 1996):

$$var(X^{(m)}) \sim m^{-1}, \text{ as } m \rightarrow \infty, \qquad \textbf{(Fml. 4.9)}$$

while LRD traffic processes can be characterised by

$$var(X^{(m)}) \sim m^{-\beta}, \text{ as } m \rightarrow \infty, \ 0 < \beta < 1. \qquad \textbf{(Fml. 4.10)}$$

The resulting plot is then fitted by simple least squares line through the resulting points in the plane. Values of the estimate $\beta$ of the asymptotic slope between -1 and 0 suggest self-similarity (LRD) while SRD is characterised by slope of -1. The estimated Hurst parameter can then be calculated by Formula 4.5. In the following 2 chapters the values of $log(var[X^{(m)}])$ are labelled as LOG10(Vm) for short.

## 4.5 Summary

In this chapter the statistical methods used to test the null hypothesis have been presented. The methods include informal and formal methods. The graphical methods as informal methods are used in conjunction with formal methods to visually clarify the results of formal methods. Two graphical methods known as EDF plot and Q-Q plot have been considered to be used. It is shown that the EDF plot with logarithmic scale is more powerful than the EDF plot with linear scale since the former can reveal any discrepancies more clearly.

The null hypothesis test results will be based on the test results of formal methods. The formal method of K-S test with V statistic has been considered as the primary statistical test since, for two samples case, it has a simple formula that relates the value of the V statistic to its significance level and it is more powerful than the well-known D statistic.

The chapter has also described shortly the second order statistic tests that will be used to test the presence of self-similarity in aggregate traffic data. The presence of self-similar behaviour or fractal nature in aggregate traffic data can be detected by estimating the Hurst parameter of aggregate traffic data. Two graphical methods known as IDC and VTP have been presented for this purpose.

# 5. Testing the Null Hypothesis: Simulation and Analysis

## 5.1 Introduction

This chapter presents experimental methodology for testing the null hypothesis and test results. In Section 5.2 we describe the experimental methodology which is applied to several experimental configurations. The configurations use different test parameters to test the sensitivity of the null hypothesis.

Section 5.3 describes traffic models and packet size distributions used in testing the null hypothesis. The two traffic models under consideration are the Poisson model and the Pareto Model. The Poisson traffic source model, with its exponentially distributed packet inter-arrival time, represents non-bursty traffic as usually used in traditional traffic modelling. The Pareto traffic source model resembles bursty traffic that is closer to real traffic data. Two different packet size distributions are used to provide two sets of user parameters. These distributions are selected to represent actual packet size distribution observed in real Ethernet traffic data.

Characteristics of the aggregate simulation data of those two models are investigated for self-similarity issues in Section 5.4. The techniques of IDC and VTP, described in the previous chapter, are utilised for this purpose. This is intended to provide a reference for comparing test results of those two traffic models.

Sections 5.5 and 5.6 present the results of testing the null hypothesis against several parameters. The results are presented in two sections in accordance with traffic models being used. These sections also provide analysis of test results. The test results are then summarised in Section 5.7.

## 5.2 Experimental Methodology

Prior to testing the null hypothesis it is necessary to develop an experimental methodology that serves as guidance for further analysis. It is also necessary to define a reference test result as a rule to which other test will be referred to for comparison. For each traffic model, the following steps are performed to test the null hypothesis:

1. Define 3 stations to represent different user parameters, i.e. users with different activity level. We define Station 1 (S1) as low-level activity station with 10 packets/second arrival rate, Station 5 (S5) as medium-level activity station with 50 packets/second arrival rate and Station 9 (S9) as high-level activity station with 100 packets/second arrival rate. The rates of these 3 stations are fixed for all configurations while other station rates are varied to obtain different utilisation levels. These varying rate stations serve as background traffic generators.

2. Set the simulation parameters for the reference configuration (see discussion below).
3. Change simulation parameters for each configuration associated with the different test parameters. The arrival rates of three reference stations are maintained.
4. Analyse test results and compare to reference results for each reference station.

The null hypothesis is tested by comparing IAT and IDT distributions for each reference station. To draw conclusions the results of other simulation configurations are compared to reference configurations. As a rule we define a 0.05 significance level to reject the null hypothesis, i.e. the null hypothesis is accepted if the significance probability from K-S test is equal or greater than 0.05. In other words, if the significance probability is less than 0.05, i.e., very small, then the null hypothesis is rejected (Press et al. 1992, pp. 620-627) since the two distributions are significantly different.

Based on discussion in Chapter 2, to isolate other factors that may affect the null hypothesis while maintaining the validity of simulation results, we define the following configuration as a reference configuration:

Star network with 30 stations, 23 meters distance between stations, large buffer to prevent overflow at each station, packet size distribution of type 1 (see the following section). This configuration is then labelled *E30SNLBP1* which can be interpreted as configuration with 30 stations with Exponential IAT (E30; we use initial P for Pareto IAT), in short network (SN), with large buffer (LB) at each station and packet size distribution of type 1 (P1). Other configurations are defined accordingly as follows (examples for Exponential IAT):

1. *E30LNLBP1*: This is for 30 stations with Exponential IAT, in long network (LN, i.e. 750 meters distance between stations), with large buffer (LB) at each station and packet size distribution of type 1 (P1). This configuration represents different network parameter, i.e. different bus length.
2. *E30SNLBP2*: This is the same as reference configuration above but differs in packet size distribution. Here we use packet size distribution of type 2 (P2) to represent different user parameters.
3. *E30SNSBP1*: This is the same as reference configuration above but differs in size of buffer. Here we use small buffer (SB) of size 4 packets to represent different protocol parameters.
4. *E50SNLBP1*: As in reference configuration but differs in number of stations. Here we use 50 stations to represent different network parameters.

Accordingly, for Pareto IAT we have *P30SNLBP1* as the reference configuration and other configurations as: *P30LNLBP1, P30SNLBP2, P30SNSBP1* and *P50SNLBP1*.

In total there are 10 different configurations to test the null hypothesis. For each configuration the simulation program is executed for 9 different utilisation levels with a different random seed at each utilisation. Each configuration is then executed 5 times with different seeds to examine the bias of the statistical results. Therefore, for each configuration, the performance results of each configuration are averaged over 5 simulation runs. As in Section 3.3.2.1, the confidence intervals for the simulation results in this chapter were derived using t distributions. In this case, we use (n-1) = 4 degrees of freedom and they came out to be very small.

For performance analysis we need to execute our simulation program for 9 different utilisation levels, i.e., 5%, 10%, 20%, 30%, 40%, 50%, 60%, 75%, and 90%. Since typical load of real Ethernet networks are well below 50% and often closer to 5% (Boggs, Jefrey & Kent 1995), for the purpose of testing the null hypothesis we analyse only the first 6 utilisation levels, i.e., 5%, 10%, 20%, 30%, 40%, 50%. These utilisation levels are labelled from 1 to 6 accordingly. Thus the simulation result with utilisation index (UI) of 4 means the result is for network utilisation of 30% and so on. To achieve these utilisation level we set the traffic intensity defined in Formula 3.4 by adjusting arrival rate of stations functioning as background traffic generator. Note that the packet arrival rates of the 3 reference stations are unchanged in each utilisation level.

The important data for testing the null hypothesis is the packet inter-arrival time (IAT) distribution and packet inter-departure time (IDT) distribution of each reference station. To obtain reliable statistical results, we define different number of IAT or IDT samples for each station. For Station 1 we collect at least 4,000 pair samples for its IAT and IDT. For Station 5 we collect 20,000 pair samples while for Station 9 we collect 40,000 pair samples for their IAT and IDT data. With this sample size and the analysis results of 5 repetitions, we believe our statistical test results are justified.

We also collected the aggregate traffic time series data for the purpose of testing the self-similarity (see Section 5.4) for each utilisation level. All data is collected after 20 seconds simulated time to remove undesired transient data. This transient interval is more important in Pareto traffic model. The transient state can be identified by the domination of a station having the first chance to acquire the channel. In steady state each station transmits packets at stable rate bound up to its packet arrival rate setting. The packet departure rate thus becomes an important parameter to verify the stability of the station under consideration. The steady state packet departure rate of any station should not differ significantly from its predefined packet arrival rate.

Our study showed that the transient interval in our simulation is actually less than 10 seconds simulated time for Pareto IAT, while for Exponential IAT 2 seconds simulated time prior to data collection is enough to remove the transient data. However for the sake of stability of results we define 20 seconds simulated time as transient interval. All data is then recorded after 20 seconds simulated time has elapsed. All simulations are then terminated after the minimum sample sizes for reference stations defined above are achieved. For the 6 level of utilisation defined above a total number of 1,000,000 departed packets is enough to terminate the simulation. On a LINUX platform in Pentium 166 machine it takes on average 5 hours to run 1 simulation (9 utilisation levels) with 5 times repetition and data analysis.

For each simulation data (IAT and IDT for each reference station) we calculate their descriptive statistics (minimum, maximum, mean, variance, standard deviation, average deviation or mean absolute deviation, coefficient of variation, skewness, kurtosis and Pearson's linear correlation coefficient). These descriptive statistics will be used to support analysis whenever necessary. The IAT and IDT data is then sorted in ascending order before being applied to the K-S test programs. Following this the aggregate time series data is then analysed for the self-similarity issues. This completes the statistical calculation part of our simulation program. All results are saved in files for further analysis.

### 5.3 Traffic Source Models and Packet Size Distributions

The two traffic models used in our simulation program are Poisson model and a 2 parameters Pareto model. The Poisson traffic model is characterised by exponential distribution of inter-arrival time. Its distribution function with scale parameter λ is (Evans, Hastings & Peacock 1993, pp. 59-62)

$$F(x) = [PX \leq x] = 1 - e^{-\lambda x} \qquad \textbf{(Fml. 5.1)}$$

To generate exponential packet inter-arrival time, we employee transformation technique (Leon-Garcia 1994, pp. 155-158) using the relationship

$$IAT = -\frac{\ln(U)}{\lambda} \qquad \textbf{(Fml. 5.2)}$$

where $U$ is uniform continuous variate between 0 and 1. The station activity level is simply its hazard rate, i.e.: (Evans, Hastings & Peacock 1993, p 59).

$$Poisson\ Packet\ Arrival\ Rate = 1/mean = \lambda \qquad \textbf{(Fml. 5.3)}$$

For Pareto traffic model, we use the so-called Pareto I distribution (Johnson, Kotz & Balakrishnan 1994, p 574). It has two parameters i.e., location parameter ω and shape parameter β. The distribution function of Pareto I is

$$F(x) = P[X \le x] = 1 - (\omega / x)^{\beta}, \quad \omega, \beta \ge 0, \quad x \ge \omega \qquad \textbf{(Fml. 5.4)}$$

As with the Poisson traffic model, to generate Pareto I IAT distribution we also use transformation method using the relationship (Evans, Hastings & Peacock 1993, p 122).

$$IAT = \omega(1 - U)^{-1/\beta} \qquad \textbf{(Fml. 5.5)}$$

where $U$ is uniform continuous variate between 0 and 1. The station activity level is defined only if its mean exists, i.e.:

$$Pareto\ Packet\ Arrival\ Rate = 1/mean = (\beta - 1) / (\omega\beta), \quad \beta > 1 \qquad \textbf{(Fml. 5.6)}$$

Based on TELNET traffic measurements, Paxson and Floyd (1994) indicated that the main body of distribution of TELNET inter-packet times can be well approximated using a Pareto distribution with a shape parameter between 0.9 and 0.95. This means the distribution has infinite mean and infinite variance. However, our simulation data showed that Pareto distribution with shape parameter β ≤ 1 resulted in very bursty traffic, unlike real traffic data. The same result has also been reported by Ryu and Lowen (1996) in their study of modelling self-similar network traffic with Fractal Renewal Point Process (RFPP). For better modelling of network traffic they suggested a value of β in the range 1 < β < 2. Since we also need a reference for comparison study between Poisson traffic model and Pareto traffic model we have set the Pareto traffic generator in our simulation with shape parameter of 1.5 for the main Pareto configurations. This setting allows us to assign the same packet arrival rate in either Poisson traffic model or Pareto traffic model since the mean of this distribution does exist but the variance is infinite. However we also execute our simulation program with different shape parameters to study the effect of this parameter on the null hypothesis. In this case we consider different shape parameters of Pareto traffic model as different user characteristics.

The main 5 simulation configurations, with the Pareto traffic model described above, are then executed with shape parameter of 1.5. One implication of this setting is that the resultant location parameter, which defines the minimum packet's IAT, (see Formula 5.6) will be greater then the minimum packet's IDT i.e., 67.2 μs (Boggs, Mogul, Kent 1995) which will affect the results of K-S test. To examine the effect of different shape parameters we ran our simulation program with reference configuration fed with different shape parameters ranging from 1.2 to 2.5. These experiments also useful to verify the self-similarity nature in aggregate traffic data for different shape parameter, i.e. the relation defined by Formula 4.8.

As explained in the previous chapter, the burstiness of aggregate traffic data can be measured by estimating its Hurst parameter that is related to the Noah intensity parameter $\alpha$ of individual traffic source (Willinger et al. 1995). The lower the value of this parameter the higher the Hurst parameter will be (see Formula 4.8). This means more bursty traffic can be generated by a source having lower $\alpha$. In Pareto traffic modelling this parameter is related directly to the shape parameter $\beta$.

For a shape parameter of 1.2 the location parameter is still in order of milliseconds. For this value the K-S test will obviously detect the discrepancy in lower tail of Pareto distribution. Therefore for complete assessment of the effect of different shape parameters we also run our simulation with the location parameter of 0.0000672 which corresponds to the shape parameters of 1.00067, 1.00337 and 1.00677 for Station 1, 5 and 9 respectively. This configuration yields very bursty traffic since the shape parameter is very closed to 1.

### 5.3.1 Packet Size Distributions

Two packet size distributions are used to represent different user parameters. As mentioned in the previous chapter, to simplify the packet generation process we have selected four packet lengths, i.e. 64 bytes, 270 bytes, 1080 bytes and 1514 bytes as valid packet length in our simulation program. The distribution of these packet sizes are set to resemble real Ethernet packet size distribution which usually consists of many short length packets, some maximum length packets and a few of intermediate size (Gusella 1990; Schoch & Hupp 1980; Crane 1981).

The two packet size distributions (labelled as P1 and P2) have basically bimodal distributions where each mode consists of two packet sizes as shown in Table 5.1. The P1 distribution has mean packet length of 250.25 bytes while the mean packet length of P2 is 426.20 bytes. We assume these two packet size distributions represent different applications, as user parameters, running in the network. During simulation execution the packet size for each arriving packet is selected randomly while preserving the overall ratio. The packet size distributions as presented in Table 5.1 are comparable to a real packet size distribution as we observed in our measurement study (see Section 6.3).

Table 5.1 Packet Size Distribution of P1 and P2

| Packet Length (bytes) | 64 | 270 | 1080 | 1514 |
|---|---|---|---|---|
| Probability of P1 | 0.80 | 0.075 | 0.025 | 0.10 |
| Probability of P2 | 0.65 | 0.10 | 0.05 | 0.20 |

## 5.4 Characteristics of Simulation Aggregate Traffic Data

This section presents the self-similarity test results performed on aggregate traffic of simulation data using IDC and VTP methods. The results will be used as reference when comparing test results of two traffic models in relation to real traffic characteristics. Since one simulation execution results in lot of data, to save space we only present typical results and describe interesting characteristics accordingly.

### 5.4.1 Characteristics of Simulation Aggregate Data with Poisson Traffic Model

Figure 5.1 depicts a typical IDC plot of aggregate traffic simulation data with Poisson traffic model. Its related typical VTP plotting is presented in Figure 5.2. It is apparent from the two figures that the aggregate traffic of simulation data generated by stations having exponential packet IAT distribution behaves like Poisson processes, i.e. having short-range dependence. In fact the estimated Hurst parameter of this data is 0.5 on average. Table 5.2 presents estimated Hurst parameter of various simulation configuration with Poisson traffic model averaged over 5 runs. The estimated Hurst parameter obtained by IDC method and VTP method is the same to 4 decimal places.

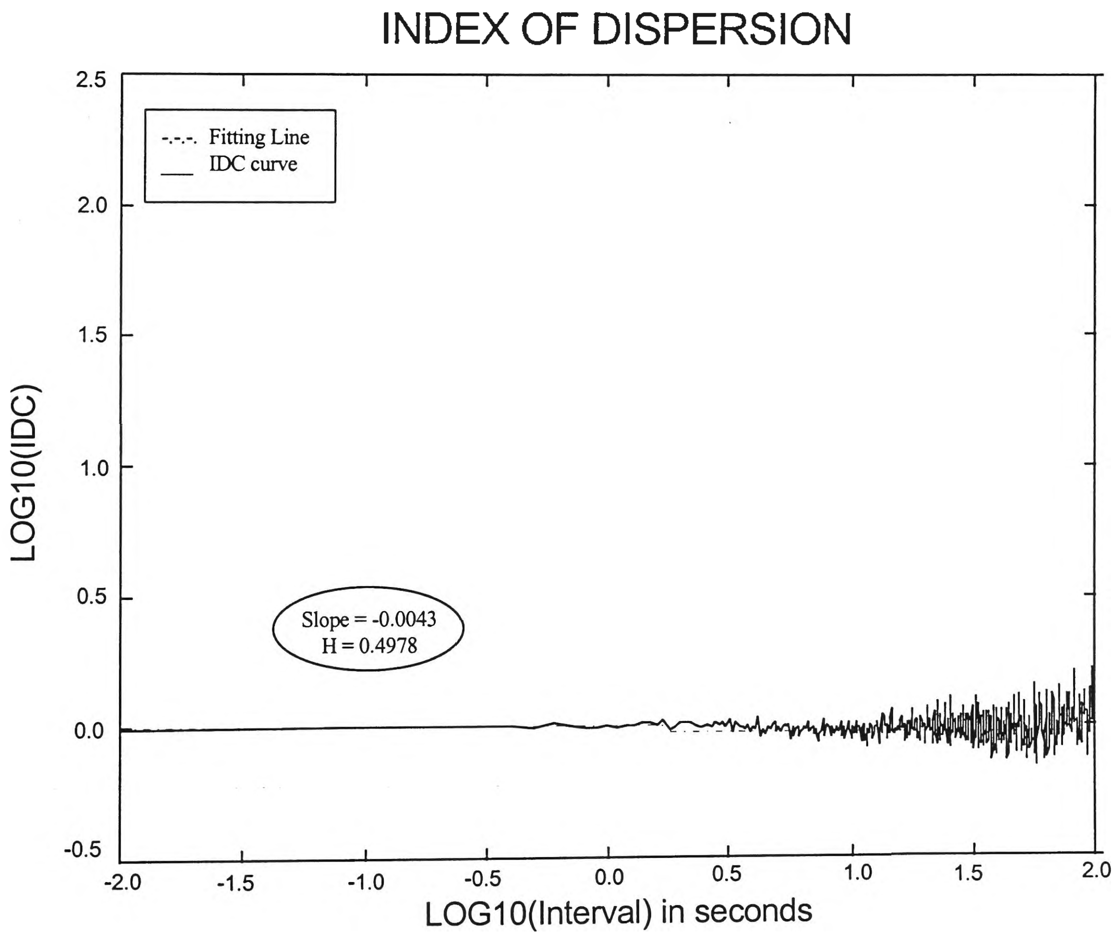


Figure 5.1 Typical IDC Plot of Simulation Data with Poisson Traffic Model

Table 5.2 Estimated H Parameters of Simulation Data with Poisson Traffic Model

| Configuration / Util. Index | *E30SNLBP1* | *E30LNLBP1* | *E30SNLBP2* | *E30SNSBP1* | *E50SNLBP1* |
|---|---|---|---|---|---|
| 1 (5%) | 0.4961 | 0.4961 | 0.5094 | 0.4919 | 0.4978 |
| 2 (10%) | 0.5079 | 0.4973 | 0.5055 | 0.5045 | 0.5024 |
| 3 (20%) | 0.5042 | 0.4968 | 0.5045 | 0.4970 | 0.5040 |
| 4 (30%) | 0.5017 | 0.4872 | 0.5089 | 0.4918 | 0.4961 |
| 5( 40%) | 0.4969 | 0.5014 | 0.5125 | 0.5000 | 0.5073 |
| 6 (50%) | 0.4958 | 0.4897 | 0.5022 | 0.4891 | 0.4855 |

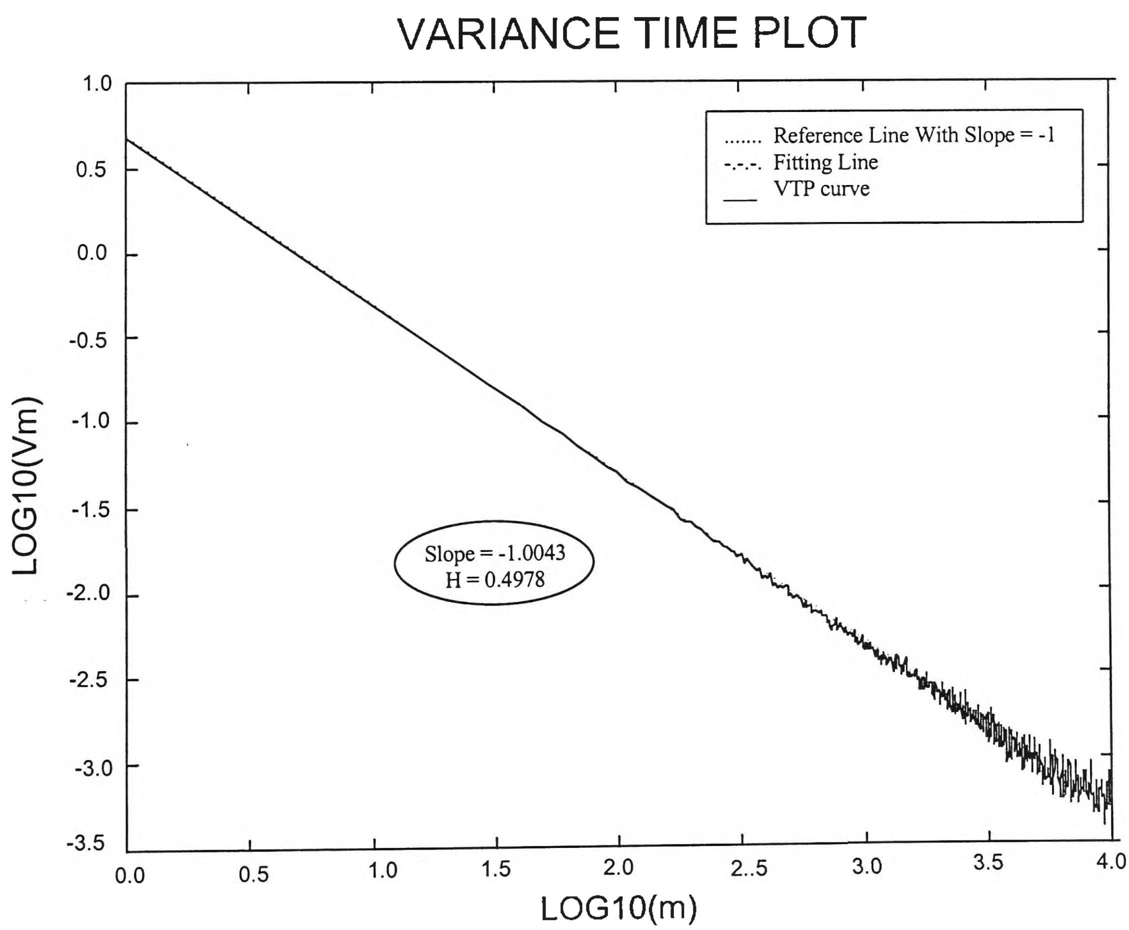


Figure 5.2 Typical VTP Plot of Simulation Data with Poisson Traffic Model

As can be seen in Table 5.3, the estimated H parameters are quite stable at value 0.5, which indicates short-range dependent characteristic.

### 5.4.2 Characteristics of Simulation Aggregate Data with Pareto Traffic Model

Figure 5.3 depicts a typical IDC plot of simulation aggregate traffic data with Pareto I traffic model. Its related typical VTP plot is presented in Figure 5.4. Both figures indicate that aggregate traffic of simulation data generated by stations having Pareto packet IAT distribution has long-range dependence property across a wide range of time scales. The plots are comparable to the same plots of real traffic data as observed in previous studies (Fowler & Leland 1991; Leland et al. 1994).

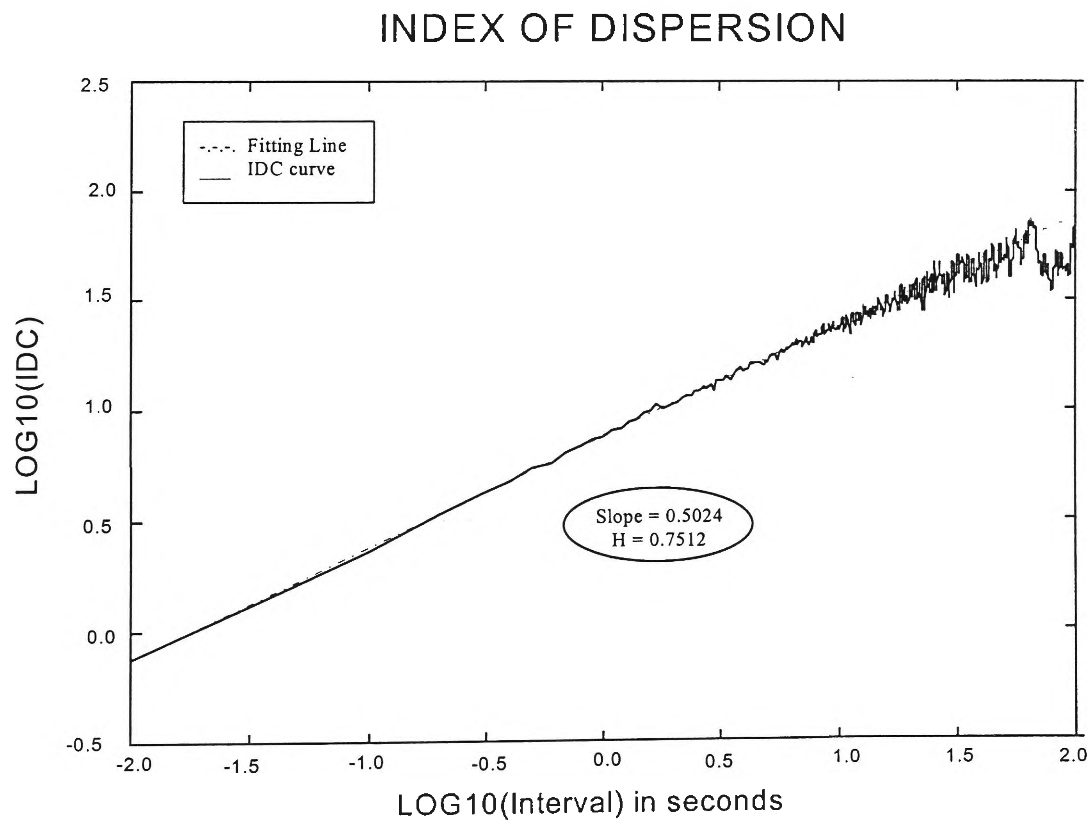


Figure 5.3 Typical IDC Plot of Simulation Data with Pareto Traffic Model

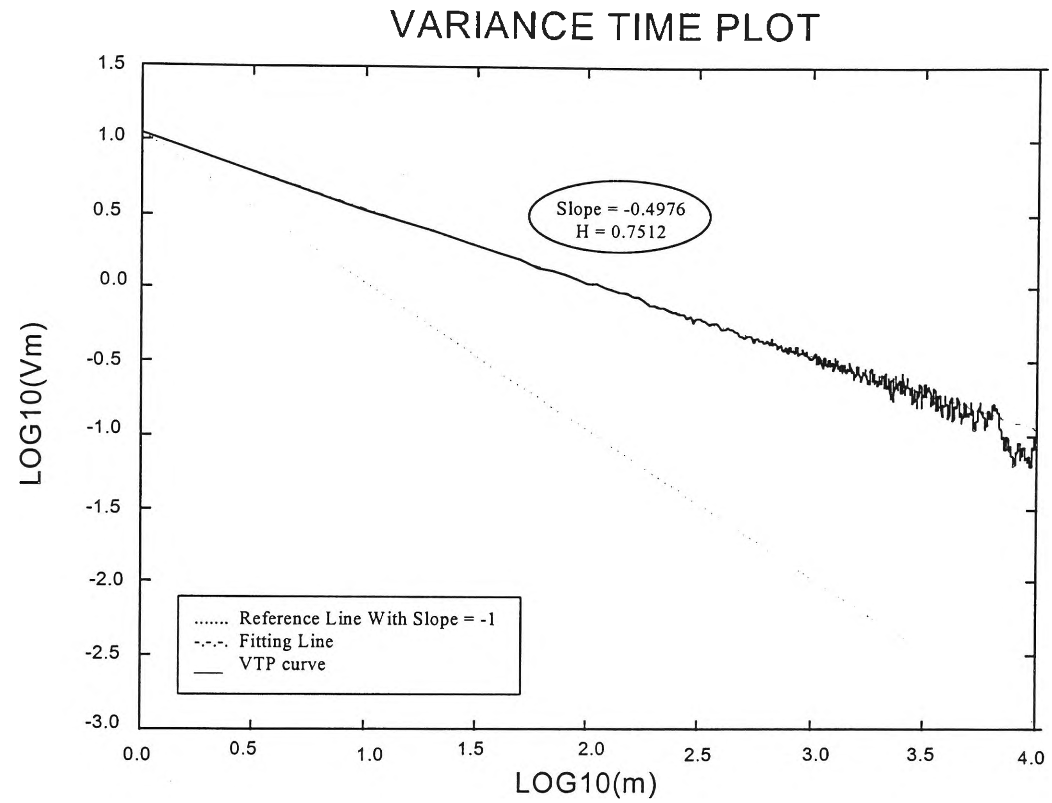


Figure 5.4 Typical VTP Plot of Simulation Data with Pareto Traffic Model

Table 5.3 presents estimated H parameters for the 5 simulation configurations fed with Pareto I IAT with shape parameter 1.5. The results are averaged over 5 simulation runs. In addition we also execute the Pareto reference configuration program (for 30 and 50 stations) with different shape parameters ranging from 1.2 to 2.5. The resultant estimated H parameters are summarised in Table 5.4.

As can be seen from Table 5.3 the estimated H parameter is not too far from 0.75 as predicted by Formula 4.8. Comparing H parameters of 30 station configuration (column 2) and 50 station configuration (column 6) and also column 3 and 4 of Table 5.4 indicate that the burstiness of aggregate traffic data is increased as number of active user increases. This is a property of second-order self-similar processes which is in contrast to typical packet traffic models considered in the literature (Georganas 1994).

Table 5.3 Estimated H Parameters of Simulation Data with Pareto Traffic Model ($\beta = 1.5$)

| *Configuration* / *Util. Index* | *P30SNLBP1* | *P30LNLBP1* | *P30SNLBP2* | *P30SNSBP1* | *P50SNLBP1* |
|---|---|---|---|---|---|
| 1 (5%) | 0.7508 | 0.7628 | 0.7826 | 0.7562 | 0.7525 |
| 2 (10%) | 0.7315 | 0.7442 | 0.7472 | 0.7432 | 0.7401 |
| 3 (20%) | 0.7545 | 0.7539 | 0.7590 | 0.7635 | 0.7763 |
| 4 (30%) | 0.7463 | 0.7565 | 0.7481 | 0.7570 | 0.7531 |
| 5( 40%) | 0.7209 | 0.7253 | 0.7502 | 0.7199 | 0.7450 |
| 6 (50%) | 0.7331 | 0.7232 | 0.7491 | 0.7324 | 0.7567 |

Table 5.4 Estimated H Parameters of Data with Pareto Traffic Model (various $\beta$)

| *Shape Parameter* / *Util. Index* | $\beta = 1.2$ *(30 Stations)* | $\beta = 1.7$ *(30 Stations)* | $\beta = 1.7$ *(50 Stations)* | $\beta = 2.1$ *(30 Stations)* | $\beta = 2.5$ *(30 Stations)* |
|---|---|---|---|---|---|
| 1 (5%) | 0.8341 | 0.6966 | 0.6985 | 0.5798 | 05048 |
| 2 (10%) | 0.8074 | 0.6919 | 0.6942 | 0.5465 | 0.4952 |
| 3 (20%) | 0.8325 | 0.7118 | 0.7368 | 0.5936 | 0.5123 |
| 4 (30%) | 0.8351 | 0.6916 | 0.7064 | 0.6154 | 0.5500 |
| 5( 40%) | 0.8031 | 0.6858 | 0.7150 | 0.6134 | 0.5504 |
| 6 (50%) | 0.8318 | 0.7037 | 0.7195 | 0.6064 | 0.5380 |

The estimated H parameters from Table 5.4 indicate the same trend as predicted by Formula 4.8, i.e. the lower the shape parameter of Pareto traffic model the higher the

H parameters. This means the Pareto traffic model with lower shape parameter will generate more bursty traffic data.

## 5.5 The Null Hypothesis Test Results for the Poisson Traffic Model

Firstly we present the performance analysis in terms of Throughput versus Offered Load and Mean Packet Delay versus Offered Load for simulation configurations fed with Poisson traffic model. They are depicted in Figure 5.5 and Figure 5.6 respectively. The results are averaged over 5 simulation runs. For this presentation we use the results of all utilisation levels ranging from 5% to 90%. As in simulation validation part the errors of the 95% confidence interval with 4 degrees of freedom are very small compared to the average value and therefore not depicted in figures.

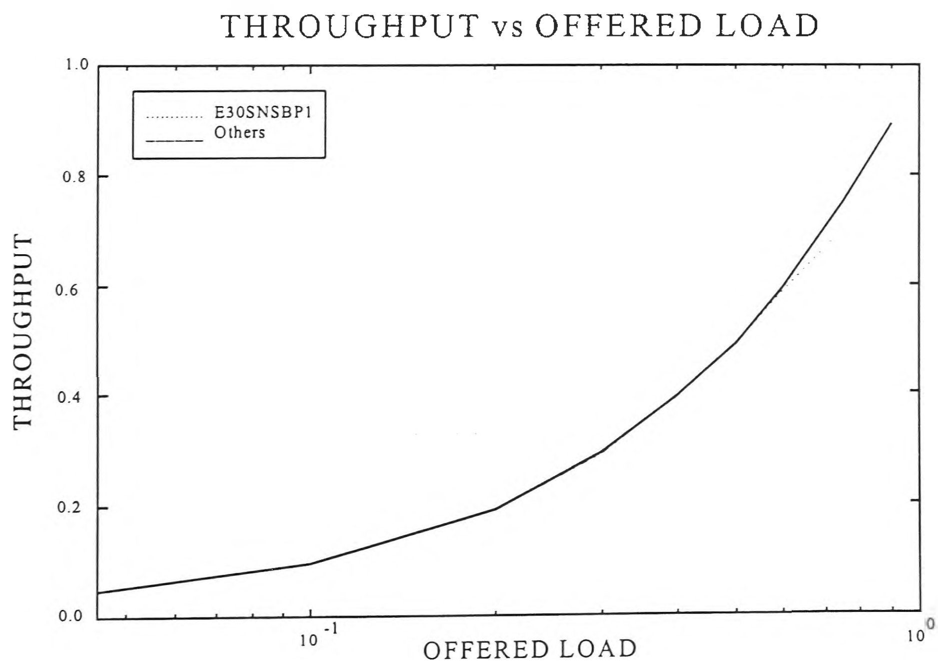


Figure 5.5 Throughput vs. Total Offered Load of Simulation Configurations with Poisson Traffic Model

As expected Figure 5.5 indicates that the performance of all configurations are about the same since we have set the same network load for all configurations. The difference is present in case where we assign small buffer to each station (E30SNSBP1). Since the buffer is small it is more likely that Ethernet protocol will reject more packets when the buffer is already full. This in turn results in lower throughput. See discussion in Chapter 2 for more detail.

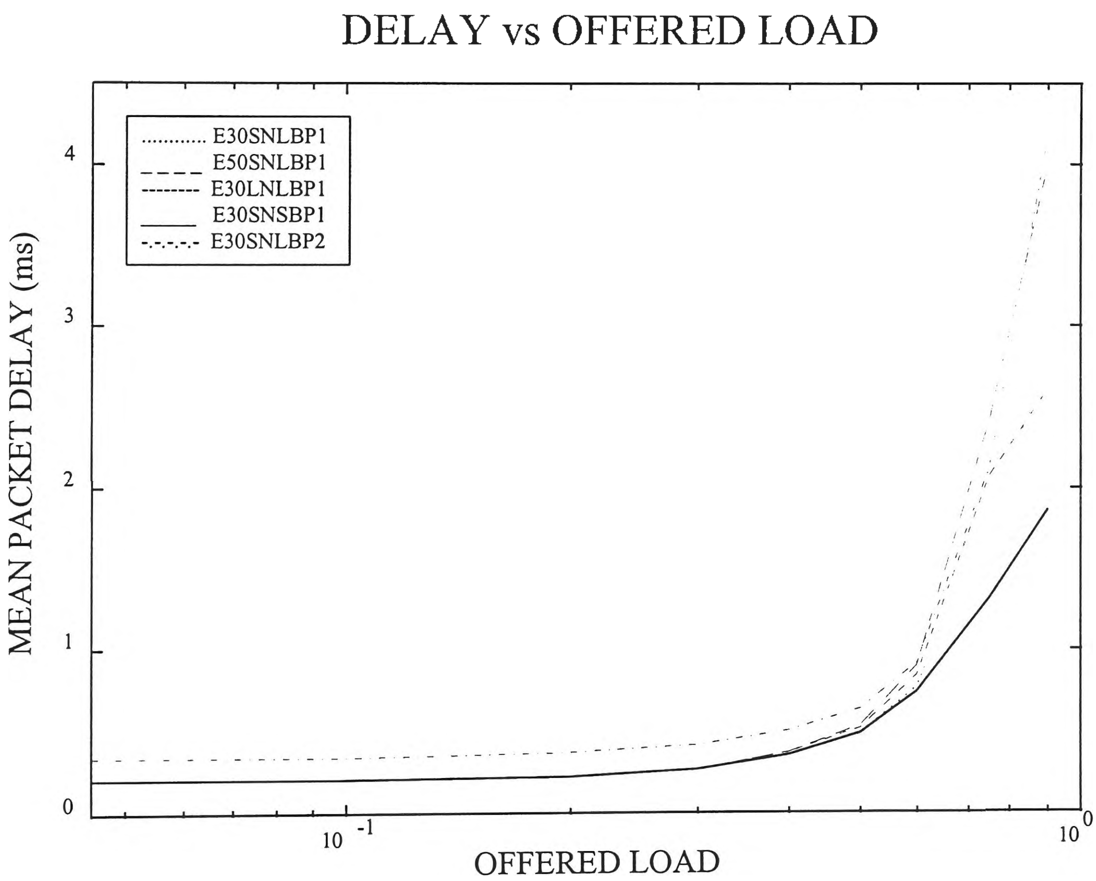


Figure 5.6 Mean Packet Delay vs. Total Offered Load of Simulation Configurations with Poisson Traffic Model

Figure 5.6 indicates that up to 50% utilisation level there are two groups of delay curves. The mean packet delay of configuration with P2 packet size distribution (E30LNLBP2) is higher than other configurations since the mean packet size also greater. Recall that mean packet delay calculation in this case takes into account the packet size as well. Above 50% utilisation level mean packet delay starts to increase exponentially. The same trend has also been reported in the previous measurement study (Gonsalves 1987).

### 5.5.1 Summary of General Statistics

From several descriptive statistics calculated in our program perhaps the most important one is the mean of IAT and IDT while other descriptive statistics do not give much more information dealing with the null hypothesis. As mentioned earlier the means of IAT and IDT distributions are related directly to station activity level as given by Formula 5.3 for Poisson traffic model or Formula 5.6 for Pareto traffic model. As a rule a pair of these statistics should not differ significantly since the Ethernet protocol should introduce delay as low as possible and be able to transmit packets as fast as possible within its limit.

To verify that the means of IAT and IDT distributions are stable we calculate the means of IAT and IDT distributions for all configurations. From 5 different runs for the same utilisation level we calculate their averages and confidence interval for 95% confidence level and 4 degrees of freedom. Since the calculation results for the 5 configurations with Poisson traffic model are similar we present only the results for the reference configuration (E30SNLBP1).

Table 5.5 presents the mean calculation for Station 1. The results are presented in their average values followed by their error separated by ± sign. These values thus give confidence limits (Sadiku & Ilyas 1995, p 76) of the statistic under consideration which are calculated with 95% confidence interval and 4 degrees of freedom. The same statistics for Station 5 and Station 9 are presented in Tables 5.6 and Table 5.7 respectively. In addition we also calculated their coefficient of variation (*CV*) and Pearson's linear correlation coefficient (r). The Pearson correlation coefficient for all pairs of IAT and IDT data are approximately 1.0, which indicates very high correlation. However, the later statistic is actually more meaningful for comparing two distributions having binormal joint probability distributions (Press et al. 1992, p 638).

Table 5.5 Descriptive Statistics of Station 1's IAT and IDT Distributions

| Utilisation | IAT | | IDT | |
|---|---|---|---|---|
| Index | Mean (sec.) | CV | Mean (sec) | CV |
| 1 (5%) | 0.10129 ± 0.00182 | 1.00238 ± 0.00535 | 0.10129 ± 0.00182 | 1.00239 ± 0.00536 |
| 2 (10%) | 0.09990 ± 0.00123 | 0.99549 ± 0.00363 | 0.09990 ± 0.00123 | 0.99547 ± 0.00362 |
| 3 (20%) | 0.09923 ± 0.00129 | 0.99675 ± 0.01826 | 0.09923 ± 0.00129 | 0.99668 ± 0.01828 |
| 4 (30%) | 0.10053 ± 0.00116 | 0.99593 ± 0.00602 | 0.10053 ± 0.00116 | 0.99589 ± 0.00610 |
| 5 (40%) | 0.09880 ± 0.00040 | 1.00114 ± 0.00664 | 0.09880 ± 0.00040 | 1.00119 ± 0.00669 |
| 6 (50%) | 0.10101 ± 0.00066 | 1.00125 ± 0.01016 | 0.10100 ± 0.00066 | 1.00123 ± 0.01014 |

Table 5.6 Descriptive Statistics of Station 5's IAT and IDT Distributions

| Utilisation | IAT | | IDT | |
|---|---|---|---|---|
| Index | Mean (sec.) | CV | Mean (sec) | CV |
| 1 (5%) | 0.02006 ± 8.9E-05 | 1.00402 ± 0.00672 | 0.02006 ± 8.9E-05 | 1.00384 ± 0.00671 |
| 2 (10%) | 0.01996 ± 0.00011 | 0.99934 ± 0.00423 | 0.01996 ± 0.00011 | 0.99915 ± 0.00422 |
| 3 (20%) | 0.01996 ± 7.9E-05 | 1.00194 ± 0.00638 | 0.01996 ± 7.9E-05 | 1.00185 ± 0.00637 |
| 4 (30%) | 0.02008 ± 0.0001 | 1.00026 ± 0.00476 | 0.02008 ± 0.0001 | 1.00032 ± 0.00470 |
| 5 (40%) | 0.02001 ± 0.00012 | 1.00070± 0.00620 | 0.02001 ± 0.00012 | 1.00100± 0.00625 |
| 6 (50%) | 0.02001 ± 6.5E-05 | 0.99860 ± 0.00839 | 0.02001 ± 6.5E-05 | 1.00098 ± 0.00828 |

Table 5.7 Descriptive Statistics of Station 9's IAT and IDT Distributions

| Utilisation | IAT | | IDT | |
|---|---|---|---|---|
| Index | Mean (sec.) | CV | Mean (sec) | CV |
| 1 (5%) | 0.01002 ± 4.8E-05 | 1.00103 ± 0.00594 | 0.01002 ± 4.8E-05 | 1.00027 ± 0.00596 |
| 2 (10%) | 0.00997 ± 3.7E-05 | 0.99648 ± 0.00255 | 0.00997 ± 3.7E-05 | 0.99586 ± 0.00253 |
| 3 (20%) | 0.01000 ± 4.0E-05 | 1.00454 ± 0.00274 | 0.01000 ± 4.0E-05 | 1.00421 ± 0.00266 |
| 4 (30%) | 0.00998 ± 3.0E-05 | 1.00235 ± 0.00562 | 0.00998 ± 3.0E-05 | 1.00266 ± 0.00570 |
| 5 (40%) | 0.01001 ± 2.5E-05 | 1.00143 ± 0.00216 | 0.01001 ± 2.5E-05 | 1.00398 ± 0.00293 |
| 6 (50%) | 0.01001 ± 1.2E-05 | 0.99879 ± 0.00309 | 0.01001 ± 1.1E-05 | 1.00687 ± 0.00342 |

As expected Table 5.5 through 5.7 indicate consistent results and small error. The means of IAT and IDT distributions of reference stations are the same up to the decimal places shown in the tables. It also indicates that packet arrival rate and packet departure rate during simulation are stable as expected. The value of coefficient of variation (CV) supports the previous calculation results of H parameters. A CV value very close to 1.0 essentially shows that the packet IDT data are still Poissonian. Since the aggregation of Poissonian traffic streams also yield Poissonian traffic (Schwartz 1987, p 30) this explain why the estimated Hurst parameters of simulation configurations with Poisson traffic model is about 0.5.

### 5.5.2 Ho's Sensitivity against Network Parameters

To test the sensitivity of the null hypothesis against network parameters we use the simulation results of configurations E30SNLBP1, E50SNLBP1 and E30LNLBP1, which represents different network parameters, i.e., 30 stations versus 50 stations and short bus versus long bus. To save space we only present the K-S test results for utilisation level at which the null hypothesis is begin to be rejected for each reference station.

It is worthy to note that all of the test results for all configurations reported in this chapter are consistent within 5 different runs and for the 6 utilisation levels. This implies that the simulation results are not bias due to different random seed used to generate packet traffic. Therefore all results are presented only in their typical values taken from one of the five simulation runs. As mentioned in previous chapter we use K-S test with V statistic to test the null hypothesis and calculate its significance probability. To provide reference we also calculate D statistics with their significance level. As a rule the test results with V statistics will be more strict since they are more sensitive than D statistics.

Table 5.8 presents K-S test results for Station 9 and 5 configuration E30SNLBP1 (reference configuration) for the utilisation level at which the null hypothesis is rejected. For other utilisation levels below this rejection level the null hypothesis can not be rejected with 0.05 significance level. For Station 1 with 10 packets/second arrival rate and Poisson traffic model, the test results for all configurations show that the null hypothesis is accepted until utilisation level 8, i.e. 75%. Hence we do not need to present the test results since we only interested to the first 6 level of utilisation.

Table 5.8 K-S Test Results for Station 9 and 5 Configuration E30SNLBP1

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 1 (5%) | 0.011 | 0.016 | 0.021 | 0.000 |
| 5 | 5 (40%) | 0.009 | 0.233 | 0.018 | 0.013 |

It is important to note that the value of D statistics or V statistics in case of 2 sample tests are not comparable to the same values of 1 sample test. In 1 sample test in which the theoretical distribution is used as reference distribution the value of D or V statistics can be obtained from statistic tables (D'Agostino & Stephens 1986, p 104-105). In a 2-sample case, as in our simulation data, we have avoided distribution's parameter estimation tasks and prefer to use direct comparison between the pair of IAT and IDT distributions. This simplifies our work and also provides more accurate test results since the IDT distributions are compared directly to their causal IAT distributions rather than to their theoretical distributions with parameters estimated from the simulation data.

Figure 5.7 presents typical EDF plot in case where the null hypothesis is accepted (Station 5 utilisation level 5) while Figure 5.8 presents typical EDF plot in case where the null hypothesis is rejected (Station 5 utilisation level 6). Both plots are selected from Station 5's IAT and IDT distributions. The upper tail of EDF (complementary EDF) presented in Figure 5.7 (accepted case) are presented in Figure 5.9 while the complementary EDF of Figure 5.8 for the rejected case are depicted in Figure 5.10.

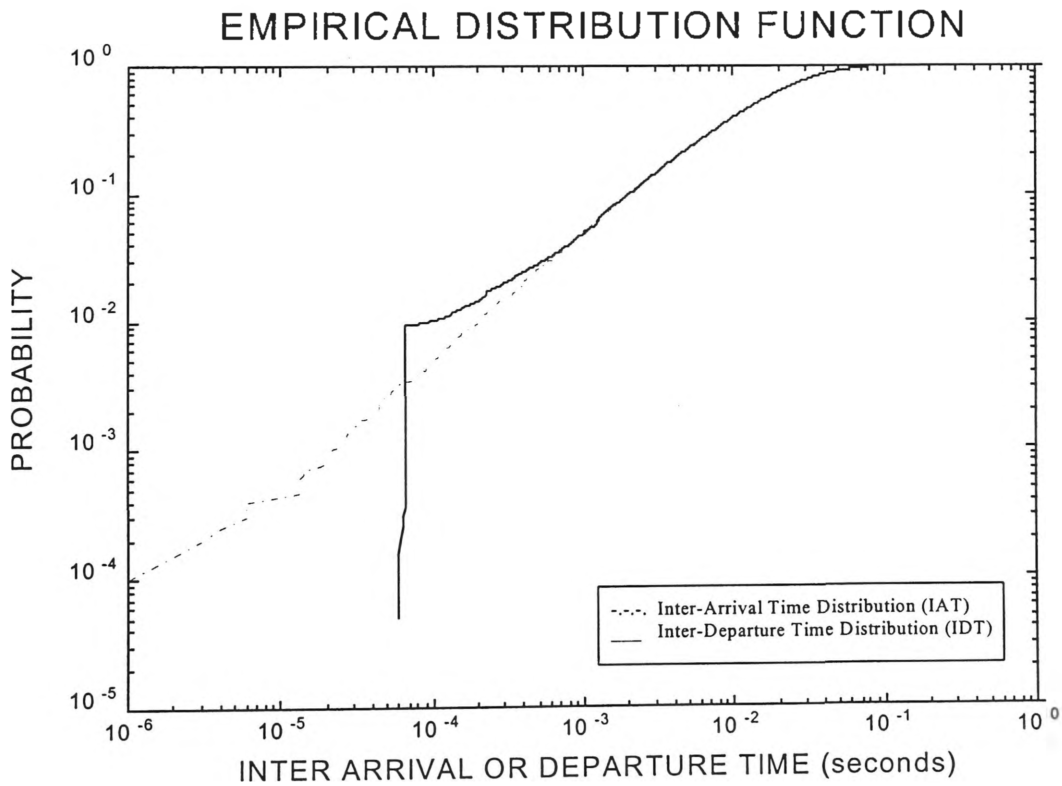


Figure 5.7 Typical EDF Plot for Accepted Case of Poisson Traffic Model Configuration

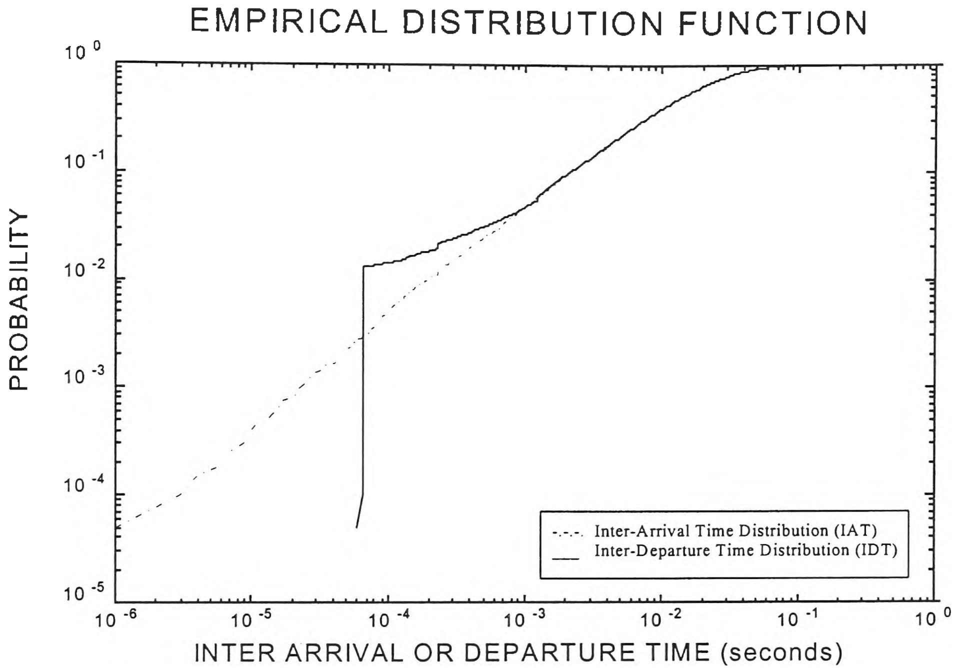


Figure 5.8 Typical EDF Plot for Rejected Case of Poisson Traffic Model Configuration

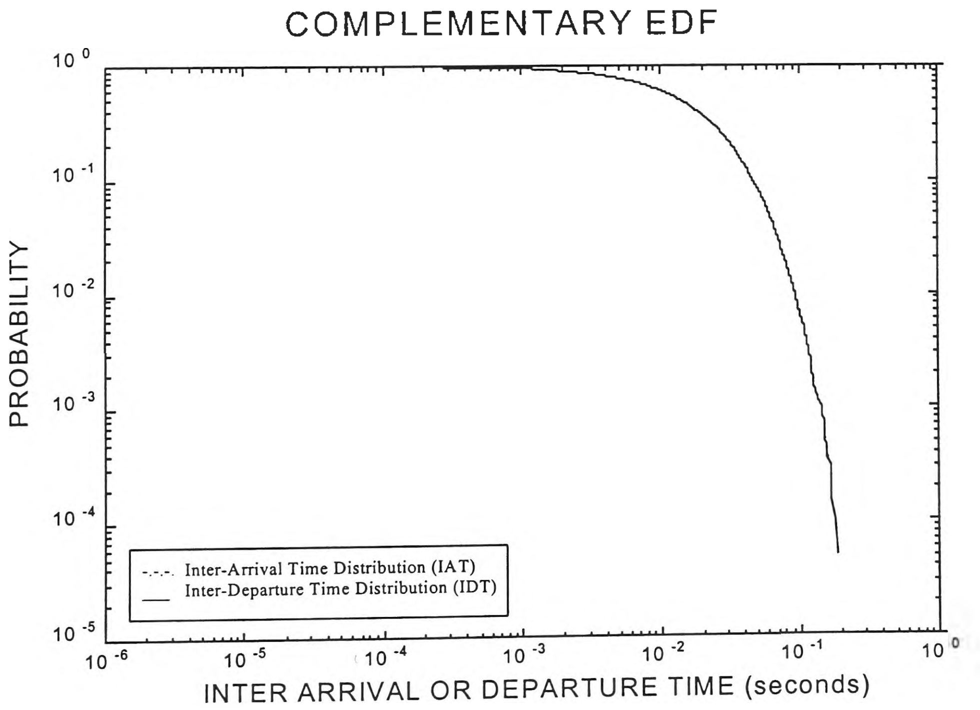


Figure 5.9 Typical Complementary EDF Plot for Accepted Case of Poisson Traffic Model Configuration

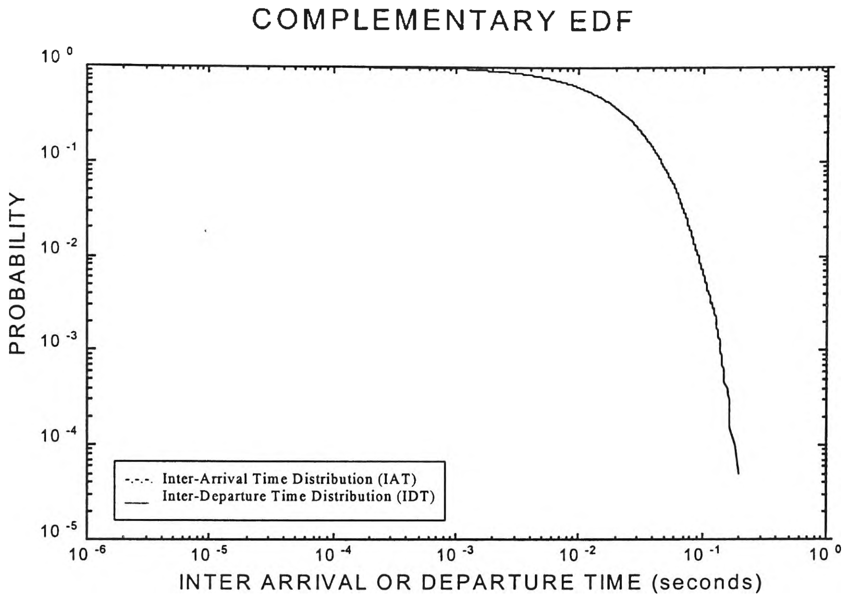


Figure 5.10 Typical Complementary EDF Plot for Rejected Case of Poisson Traffic Model Configuration

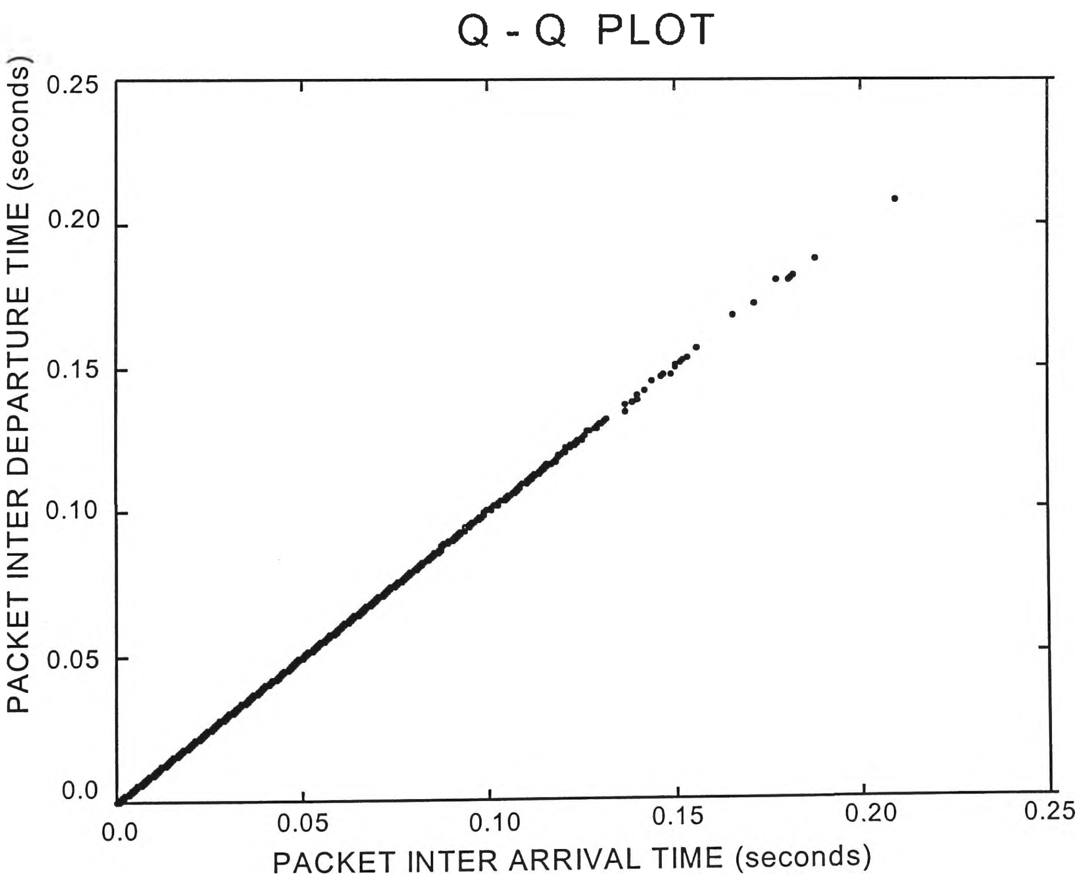


Figure 5.11 Typical Q-Q Plot for Rejected Case of Poisson Traffic Model Configuration

As can be seen from those figures, the difference is mostly present in lower tail while the upper tail of both distributions are indistinguishable. As depicted in Figure 5.8 one factor the cause the difference in lower tail is the minimum value of IDT distribution that is greater than the minimum value of IAT distribution. The minimum value of IAT distribution in case of Poisson traffic model can be as low as 0 while the minimum value of IDT distribution is bounded to 67.2 μs (minimum inter-packet departure time from the same station). In this case K-S test thus detects the difference in the lower tail and rejects the null hypothesis. The plots for other stations in other configurations look similar and therefore have not been presented.

Typical Q-Q plot for the EDF of Figure 5.8 are presented in Figure 5.11. Analysis based on Figure 5.11 will falsely conclude that both distributions are the same since the Q-Q plot results in straight line and starts from origin. The results of K-S test can then be well verified by means of EDF plot rather then using Q-Q plot.

Table 5.9 presents K-S test results of configuration E50SNLBP1 for Stations 9 and 5 while K-S test results of configuration E30LNLBP1 are presented in Table 5.10.

Table 5.9 K-S Test Results for Station 9 and 5 Configuration E50SNLBP1

| Reference Station | Utilisation Level of Rejection | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| | | D Statistic | Probability | V Statistic | Probability |
| 9 | 1 (5%) | 0.011 | 0.018 | 0.022 | 0.000 |
| 5 | 5 (40%) | 0.010 | 0.193 | 0.021 | 0.006 |

Table 5.10 K-S Test Results for Station 9 and 5 Configuration E30LNLBP1

| Reference Station | Utilisation Level of Rejection | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| | | D Statistic | Probability | V Statistic | Probability |
| 9 | 1 (5%) | 0.010 | 0.024 | 0.023 | 0.000 |
| 5 | 5 (40%) | 0.010 | 0.224 | 0.019 | 0.012 |

Since the K-S test results from these 3 configurations show the same rejection level of utilisation for Station 9 and 5 we conclude that for Poisson traffic model the null hypothesis is not sensitive to the network parameters under consideration.

### 5.5.3 Ho's Sensitivity against Protocol Parameter

To test the sensitivity of the null hypothesis we compare the results of configuration E30SNLBP1 to the results of configuration E30SNSBP1, which represent different packet buffers size. Table 5.11 present K-S test results for Stations 9 and 5 of configuration E30SNSBP1 for the utilisation level at which the null hypothesis is rejected.

Table 5.11 K-S Test Results for Station 9 and 5 Configuration E30SNSBP1

| Reference Station | Utilisation Level of Rejection | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| | | D Statistic | Probability | V Statistic | Probability |
| 9 | 1 (5%) | 0.011 | 0.016 | 0.022 | 0.000 |
| 5 | 5 (40%) | 0.009 | 0.242 | 0.021 | 0.005 |

Comparing Tables 5.11 to Tables 5.8 we conclude that for Poisson traffic model the null hypothesis is not sensitive to the protocol parameter under consideration.

### 5.5.4 Ho's Sensitivity against User Parameters

The first test results of the null hypothesis against user parameters are presented in Table 5.8 for reference configuration. The results indicated that the null hypothesis is sensitive to the user activity level. The higher the user activity level the more likely the null hypothesis is to be rejected. As packet arrival rate increases we may expect more packets will be queued before the station has chance to acquire the channel. Once the channel is acquired the queued packet will be transmitted continuously until the buffer is empty or a collision is detected. This will cause a shift in IDT distribution compared to IAT distribution as depicted in Figure 5.8.

Interesting test results against user parameter is presented in Table 5.12. The results indicated that the null hypothesis is sensitive to the packet size distribution. Compared to the results in Table 5.8 we can see that the null hypothesis for Station 5 is now rejected at utilisation level of 5%. It is quite reasonable since longer packets will have a greater effect on packet IDT as compared to shorter packets.

Table 5.12 K-S Test Results for Station 9 and 5 Configuration E30SNLBP2

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 1 (5%) | 0.022 | 0.000 | 0.044 | 0.000 |
| 5 | 1 (5%) | 0.012 | 0.098 | 0.023 | 0.001 |

## 5.6 The Null Hypothesis Test Results for the Pareto Traffic Model

As with the Poisson traffic model configurations we present first the performance analysis of simulation configuration fed with Pareto traffic model. Figure 5.12 depicts Throughput vs. Offered Load of simulation configurations fed with Pareto traffic model while Figure 5.13 presents Mean Packet Delay versus Offered Load for all simulation configurations with the Pareto traffic model. Both figures indicate that the performance of the simulation with the Pareto traffic model is essentially the same to that of the Poisson traffic model, as shown in Figures 5.5 and 5.6.

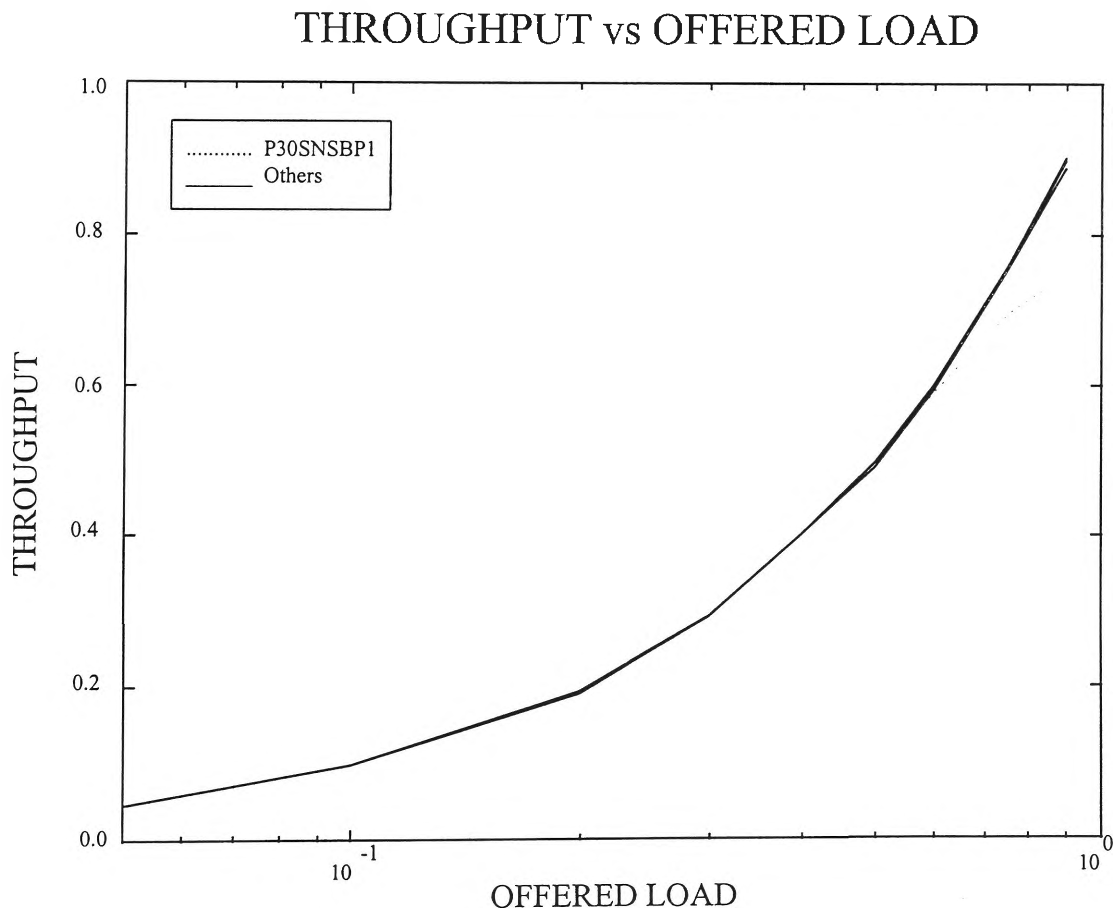


Figure 5.12 Throughput vs. Total Offered Load of Simulation Configurations with Pareto Traffic Model

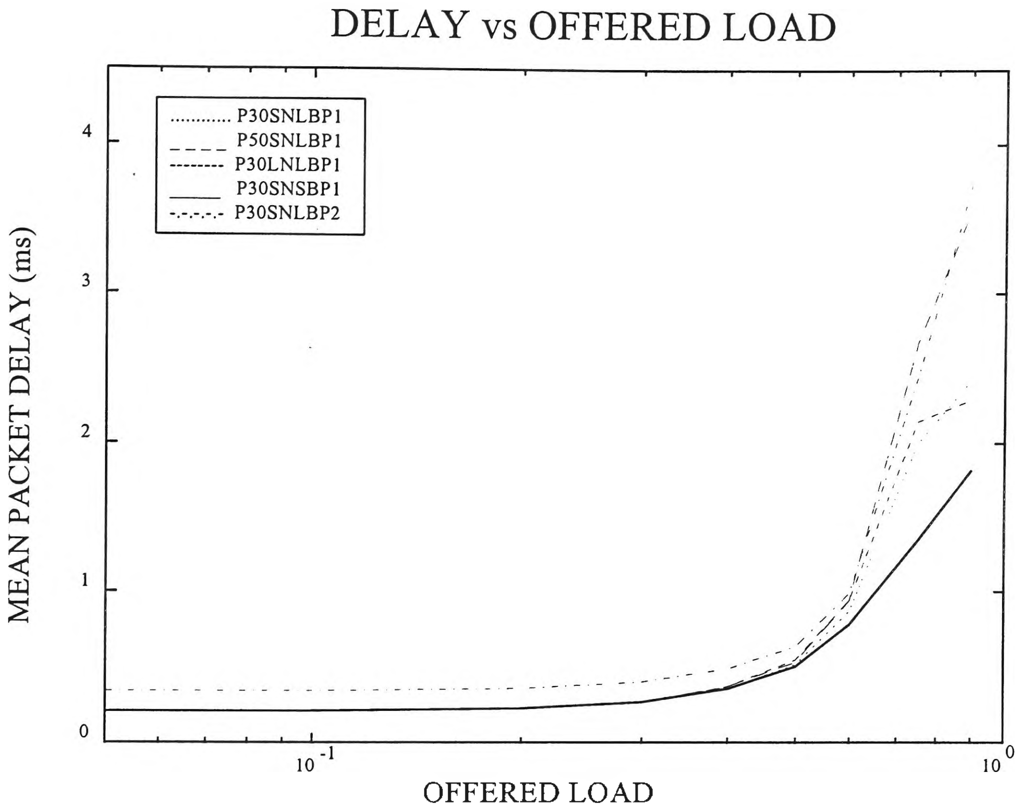


Figure 5.13 Mean Packet Delay vs. Total Offered Load of Simulation Configurations with Pareto Traffic Model

Since the test results of the Pareto traffic model will depend on the shape parameter, the test results presented in this section serve as complementary results for bursty traffic model. We first present the results for the 5 configurations fed with Pareto traffic model with shape parameter 1.5. To draw conclusion the results are then verified by the results of reference configurations fed with different shape parameters.

The K-S test results of Station 1 for all configurations with Pareto traffic model having shape parameter 1.5 show that the null hypothesis for Station 1 can not be rejected with 0.05 significance level up to utilisation level 7 (60%). Hence, we present only the K-S test results for Station 9 and Station 5 for the utilisation level at which the null hypothesis is rejected.

### 5.6.1 Summary of General Statistics

As with the Poisson traffic model we examine the calculated mean of the IAT and IDT distributions. Since theoretical variance for Pareto distributions with shape parameter less than 2 are infinite, other descriptive statistics are basically undefined. Table 5.13 through 5.15 present the mean of the IAT and IDT distributions of each reference station for reference configuration P30SNLBP1 calculated with 95% confidence interval with 4 degrees of freedom. For other configurations with this shape parameter the results are similar. As in the results of Poisson traffic model, the Pearson linear correlation coefficient for each pair of IAT and IDT data are 1.

Table 5.13 Descriptive Statistics of Station 1's IAT and IDT Distributions

| Utilisation | IAT | IDT |
|---|---|---|
| Index | Mean (sec.) | Mean (sec) |
| 1 | 0.102563 ± 0.006962 | 0.102563 ± 0.006962 |
| 2 | 0.096766 ± 0.002808 | 0.096766 ± 0.002808 |
| 3 | 0.095283 ± 0.003023 | 0.095283 ± 0.003023 |
| 4 | 0.099942 ± 0.005292 | 0.099942 ± 0.005292 |
| 5 | 0.096214 ± 0.001428 | 0.096214 ± 0.001428 |
| 6 | 0.102457 ± 0.003891 | 0.102457 ± 0.003892 |

Table 5.14 Descriptive Statistics of Station 5's IAT and IDT Distributions

| **Utilisation** | **IAT** | **IDT** |
|---|---|---|
| **Index** | **Mean (sec.)** | **Mean (sec)** |
| 1 | 0.019824 ± 0.000452 | 0.019824± 0.000452 |
| 2 | 0.019611 ± 0.000618 | 0.019611 ± 0.000618 |
| 3 | 0.020023 ± 0.000264 | 0.020023 ± 0.000264 |
| 4 | 0.019577 ± 0.000348 | 0.019577 ± 0.000348 |
| 5 | 0.019577 ± 0.000348 | 0.019577 ± 0.000348 |
| 6 | 0.019452 ± 0.000413 | 0.019452 ± 0.000413 |

Table 5.15 Descriptive Statistics of Station 9's IAT and IDT Distributions

| **Utilisation** | **IAT** | **IDT** |
|---|---|---|
| **Index** | **Mean (sec.)** | **Mean (sec)** |
| 1 (5%) | 0.009915 ± 0.000142 | 0.009915 ± 0.000142 |
| 2 (10%) | 0.009650 ± 8.98E-05 | 0.009650 ± 8.98E-05 |
| 3 (20%) | 0.010071 ± 0.000120 | 0.010071 ± 0.000120 |
| 4 (30%) | 0.009897 ± 0.000234 | 0.009897 ± 0.000234 |
| 5 (40%) | 0.009965 ± 0.000207 | 0.009965 ± 0.000207 |
| 6 (50%) | 0.009809 ± 0.000105 | 0.009809 ± 0.000105 |

As expected the means of the IAT distribution for the reference stations, which are related to the packet arrival rate, are stable. This means the simulation is in its steady state when data is collected. The means of IDT distribution also indicate that packet departure rate is the same as its packet arrival rate which is expected under normal conditions.

### 5.6.2 Ho's Sensitivity against Network Parameters

Table 5.16 presents K-S test results of Stations 9 and 5 of configuration P30SNLBP1 (reference configuration) for the utilisation level at which the null hypothesis is rejected. For other utilisation levels below this rejection level the null hypothesis can not be rejected with 0.05 significance level. As mentioned earlier the K-S test results of Station 1 for other configurations show consistently that the null hypothesis can not be rejected up to utilisation level 7 (60%) and hence the K-S test reported here do not include the results of Station 1.

Table 5.16 K-S Test Results for Station 9 and 5 Configuration P30SNLBP1

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 2 (10%) | 0.011 | 0.025 | 0.021 | 0.000 |
| 5 | 3 (20%) | 0.015 | 0.013 | 0.318 | 0.000 |

Typical EDF and ECDF plot for the accepted cases are presented in Figures 5.14 and 5.15 respectively while Figures 5.16 and Figures 5.17 depict the same plot for the rejected case. These examples are taken from Station 5's IAT and IDT distribution, i.e. utilisation level 2 for accepted case and utilisation level 3 for the rejected case. Both figures indicated the different in the minimum value of IAT and IDT distributions is one causal factor that leads to the rejection of the null hypothesis. As with the Poisson traffic model the ECDF plots of both cases are indistinguishable. Since the same plots for other configurations look similar they are not presented.

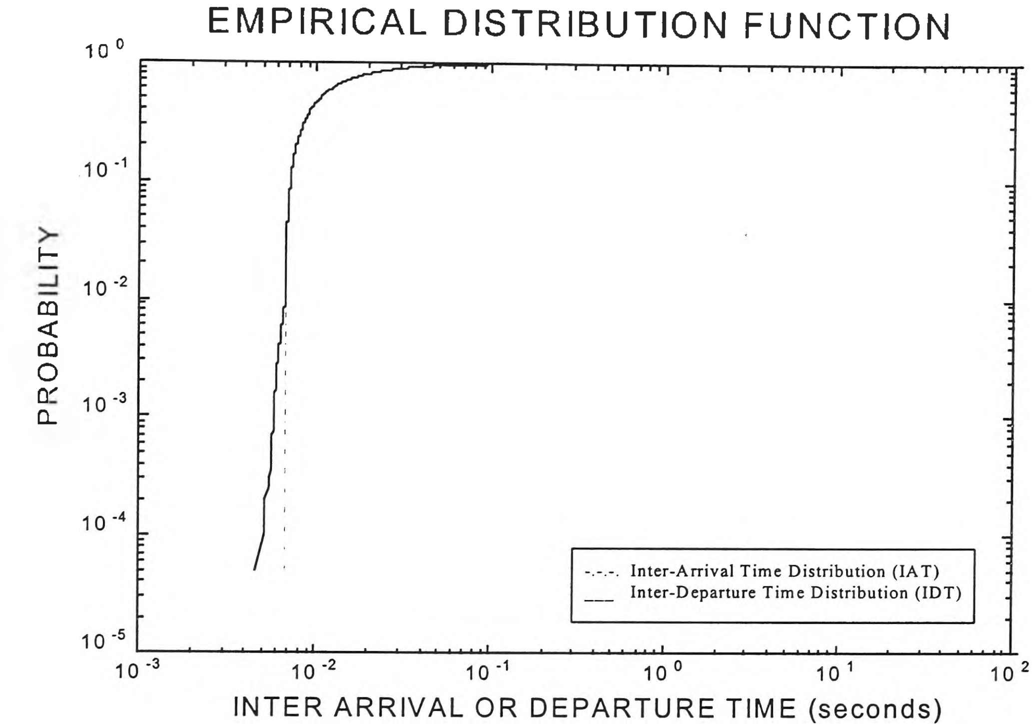


Figure 5.14 Typical EDF Plot for Accepted Case of Pareto Traffic Model Configuration

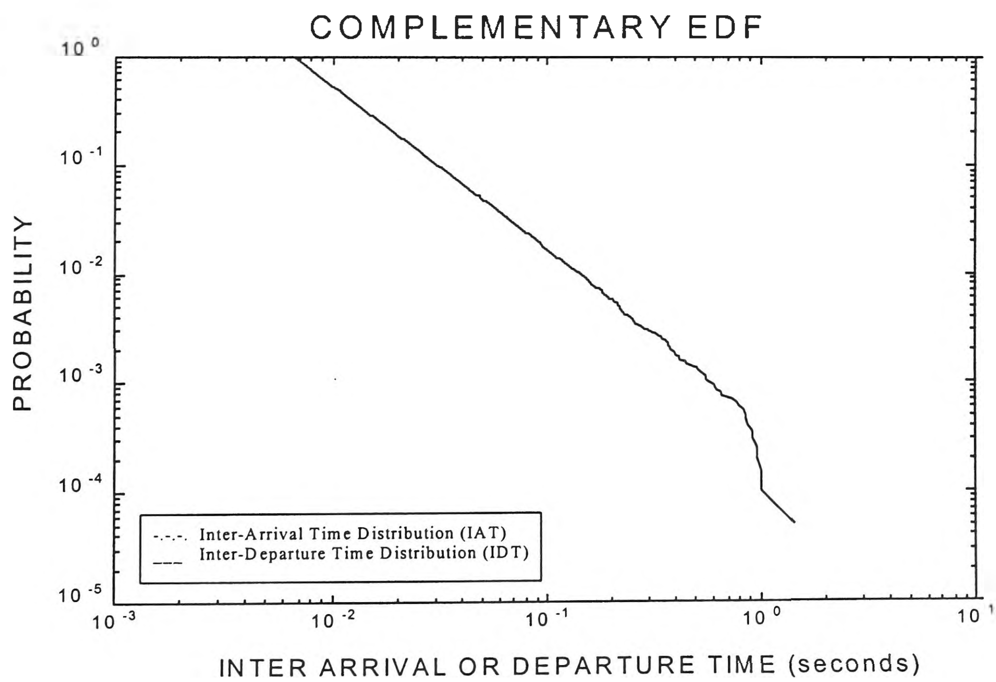


Figure 5.15 Typical Complementary EDF Plot for Accepted Case of Pareto Traffic Model Configuration

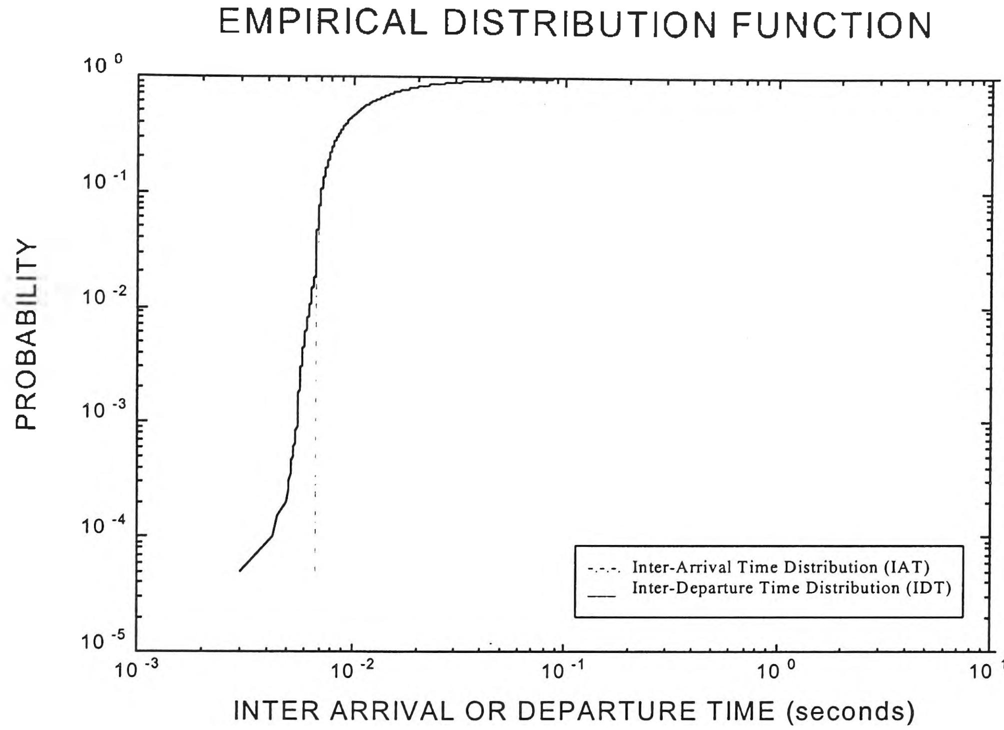


Figure 5.16 Typical EDF Plot for Rejected Case of Pareto Traffic Model Configuration

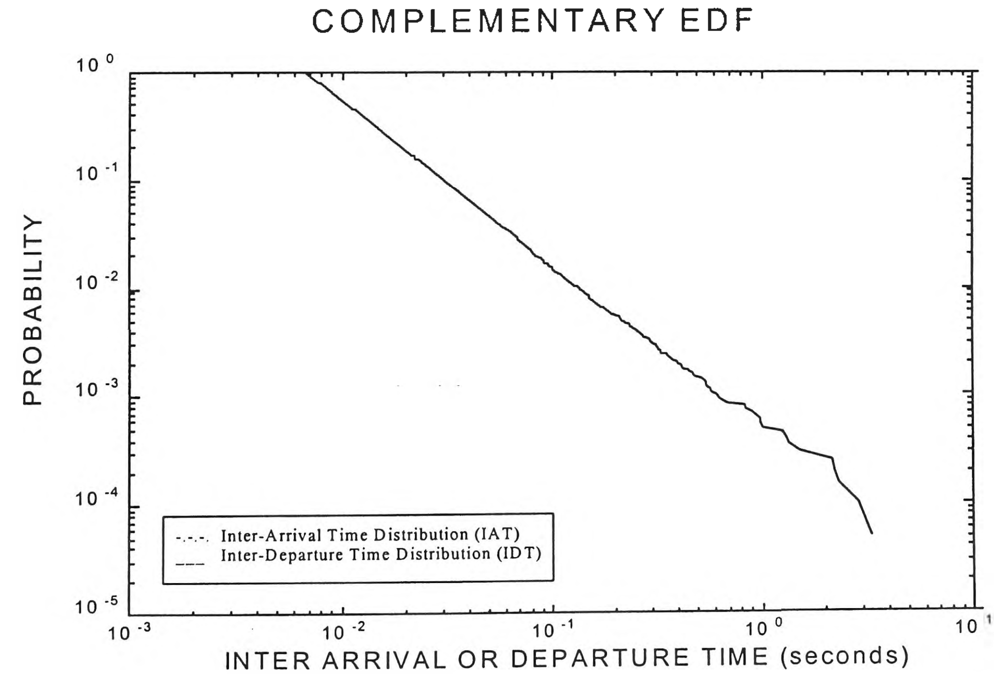


Figure 5.17 Typical Complementary EDF Plot for Rejected Case of Pareto Traffic Model Configuration

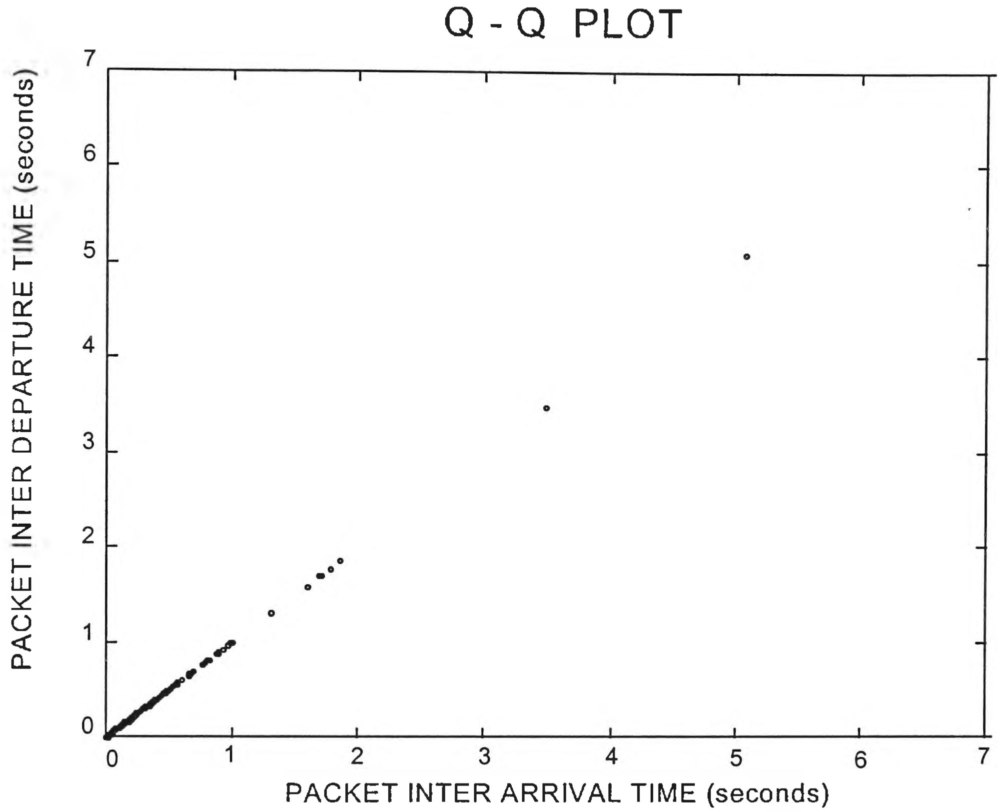


Figure 5.18 Typical Q-Q Plot for Rejected Case of Pareto Traffic Model Configuration

As with the Poisson traffic model the Q-Q plot for the rejected case of the Pareto traffic model also depicts a straight line. It is depicted in Figure 5.18. Compared to Figure 5.11, Figure 5.18 indicates the presence of clustering in lower IAT or IDT value. In the Poisson traffic model the IAT and IDT distribution is distributed more uniformly within their range but is not the case in the Pareto traffic model. This difference will affect K-S test results as shown in the subsequent K-S test results.

The following two tables present the K-S test results at which the null hypothesis is rejected for Stations 9 and 5 of configurations P50SNLBP1 and P30LNLBP1 respectively.

Table 5.17 K-S Test Results for Station 9 and 5 Configuration P50SNLBP1

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 2 (10%) | 0.012 | 0.009 | 0.023 | 0.000 |
| 5 | 3 (20%) | 0.015 | 0.019 | 0.031 | 0.000 |

Table 5.18 K-S Test Results for Station 9 and 5 Configuration P30LNLBP1

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 2 (10%) | 0.010 | 0.025 | 0.021 | 0.000 |
| 5 | 3 (20%) | 0.016 | 0.006 | 0.034 | 0.000 |

Comparing the K-S test results shown in Table 5.16 through 5.18 we conclude that for Pareto traffic model the null hypothesis is not sensitive to the network parameters under consideration. As with the Poisson traffic model, the null hypothesis is sensitive to the user parameter of activity level. However, for the Pareto traffic model the K-S test results for Station 9 is not rejected at 5 % utilisation level while in the Poisson traffic model the null hypothesis for station with high activity level is rejected at 5% utilisation level. Our studies show that at 5% utilisation level in case of the Pareto traffic model, Station 9 is the station with the highest activity level. This station tends to dominate packet transmission. Given that the main part of Pareto packet IAT is clustered at low values this in turn will introduce lower delay compared to the case with the Poisson traffic model. The packets are more likely to be transmitted as soon as they arrive after short delay. The IDT distribution thus does not differ significantly with the IAT distribution.

In case with the Poisson traffic model, the packet's IAT distributions are distributed more uniformly within its range. Any disturbance to IDT distribution, caused by different packet size for instance, will affect IDT distribution significantly as shown by the results of E30SNLBP2 configuration. The IDT distribution of Poisson traffic model thus more sensitive compared to the IDT distribution of Pareto traffic model.

### 5.6.3 Ho's Sensitivity against Protocol Parameter

Table 5.19 presents the K-S test results at which the null hypothesis is rejected for Stations 9 and 5 of configuration P30SNSBP1. This configuration represents different protocol parameter of buffer size. Each station is assigned a small buffer of 4 packets. Compared to reference configuration, this configuration caused more packets to be rejected during simulation runs. However, comparing Table 5.19 to Table 5.16 we conclude that for the Pareto traffic model the null hypothesis is not sensitive to the protocol parameter under consideration.

Table 5.19 K-S Test Results for Station 9 and 5 Configuration P30SNSBP1

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 2 (10%) | 0.012 | 0.006 | 0.024 | 0.000 |
| 5 | 3 (20%) | 0.017 | 0.006 | 0.032 | 0.000 |

### 5.6.4 Ho's Sensitivity against User Parameters

The following table present K-S test results at which the null hypothesis is rejected for Stations 9 and 5 of configuration P30SNLBP2.

Table 5.20 K-S Test Results for Station 9 and 5 Configuration P30SNLBP2

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 2 (10%) | 0.011 | 0.017 | 0.022 | 0.000 |
| 5 | 3 (20%) | 0.018 | 0.004 | 0.035 | 0.000 |

Comparing Table 5.20 to Table 5.16 we conclude that for the Pareto traffic model the null hypothesis is not sensitive to the packet size distribution. This is not the case for the Poisson traffic model where the null hypothesis is sensitive to packet size distribution. As explained earlier it is the nature of the Pareto distribution which tends to cluster at low values that makes the null hypothesis insensitive to the packet size distributions.

### 5.6.5 Ho's Sensitivity against Pareto Shape Parameters

In case of Pareto packet IAT the user activity level depend on the shape parameter of distribution. Calculation of Hurst parameters described earlier indicates that a Pareto traffic source with different shape parameters resulted in different burstiness of traffic data. It is interesting to investigate the effect of this parameter on the sensitivity of the null hypothesis. To examine this the reference configuration with Pareto traffic model has been simulated with varying shape parameters. The K-S test results for all reference stations are presented in Tables 5.21 through 5.25 that correspond to the shape parameters of 1.2, 1.5, 1.7, 2.1 and 2.5 respectively.

Table 5.21 K-S Test Results for Configuration P30SNLBP1 with $\beta = 1.2$

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 1 (5%) | 0.007 | 0.232 | 0.015 | 0.006 |
| 5 | 2 (10%) | 0.013 | 0.066 | 0.026 | 0.000 |
| 1 | 5 (40%) | 0.023 | 0.2271 | 0.046 | 0.005 |

Table 5.22 K-S Test Results for Configuration P30SNLBP1 with $\beta = 1.5$

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 2 (10%) | 0.011 | 0.025 | 0.021 | 0.000 |
| 5 | 3 (20%) | 0.015 | 0.013 | 0.318 | 0.000 |
| 1 | 7 (60%) | 0.025 | 0.204 | 0.050 | 0.004 |

Table 5.23 K-S Test Results for Configuration P30SNLBP1 with $\beta = 1.7$

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 2 (10%) | 0.009 | 0.039 | 0.019 | 0.000 |
| 5 | 3 (20%) | 0.015 | 0.013 | 0.318 | 0.000 |
| 1 | 7 (60%) | 0.023 | 0.359 | 0.046 | 0.026 |

Table 5.24 K-S Test Results for Configuration P30SNLBP1 with $\beta = 2.1$

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 2 (10%) | 0.011 | 0.013 | 0.023 | 0.000 |
| 5 | 3 (20%) | 0.015 | 0.016 | 0.030 | 0.000 |
| 1 | 8 (75%) | 0.033 | 0.105 | 0.065 | 0.000 |

Table 5.25 K-S Test Results for Configuration P30SNLBP1 with $\beta = 2.5$

| Reference | Utilisation Level | K-S Test Result | | K-S Test Result | |
|---|---|---|---|---|---|
| Station | of Rejection | D Statistic | Probability | V Statistic | Probability |
| 9 | 2 (10%) | 0.012 | 0.005 | 0.024 | 0.000 |
| 5 | 3 (20%) | 0.016 | 0.012 | 0.310 | 0.000 |
| 1 | 8 (75%) | 0.030 | 0.143 | 0.060 | 0.001 |

Comparing the K-S test results presented in Table 5.21 through Table 5.25 we conclude that the null hypothesis is sensitive to the shape parameter of Pareto distribution. The lower the shape parameter the more likely the null hypothesis to be rejected since the aggregate traffic data is more bursty.

Typical discrepancy between the IAT and IDT distributions of Pareto traffic model that lead to the rejection of the null hypothesis is depicted in Figure 5.16. The effect of location parameter on the K-S test results is apparent from Figure 5.16. The minimum value of IAT distribution for the Pareto traffic model is bounded by the value of its location parameter while the minimum value of IDT distribution is bounded by the minimum packet inter-departure time. To verify this condition we also run our simulation for the case in which the location parameter of Pareto distribution is set to be 67.2 μs. The results are presented in Figure 5.19 and 5.20. The K-S test results, in the case depicted in Figure 5.19 and 5.20, show the rejection of the null hypothesis of all reference stations for all configurations at all utilisation levels. This is due to the very bursty nature of the generated aggregate traffic data. The Q-Q plot for this case is shown in Figure 5.21. In this case the Q-Q plot can clearly indicate that the IAT and IDT distributions are not statistically the same.

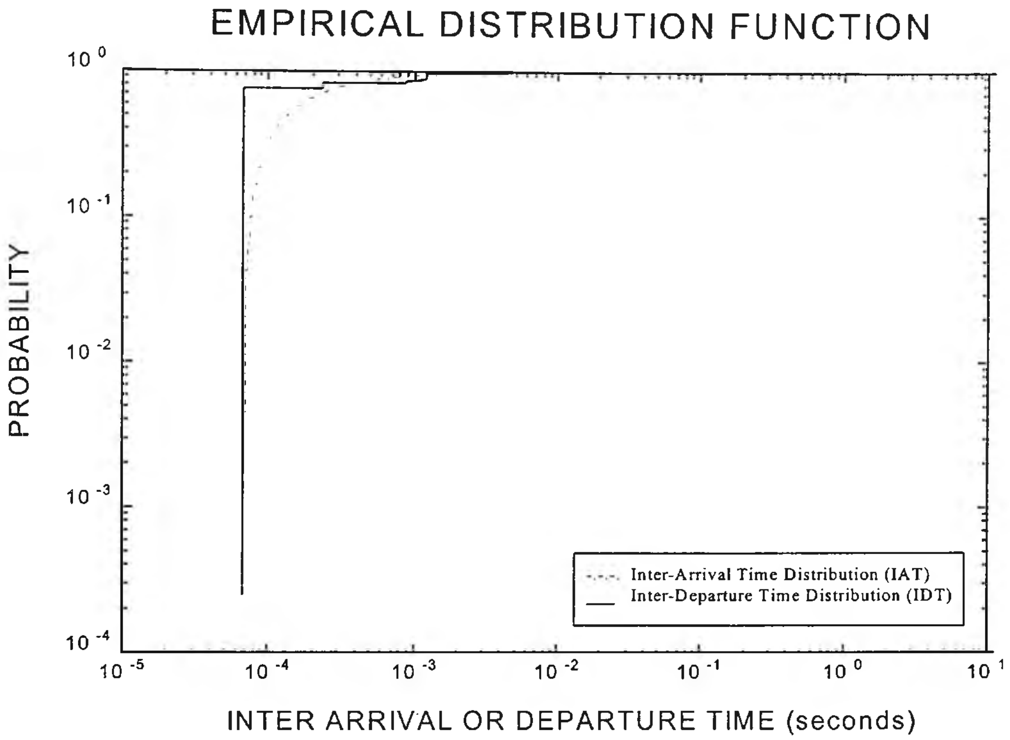


Figure 5.19 EDF Plot of P30SNLBP1 with $\omega = 0.0000672$

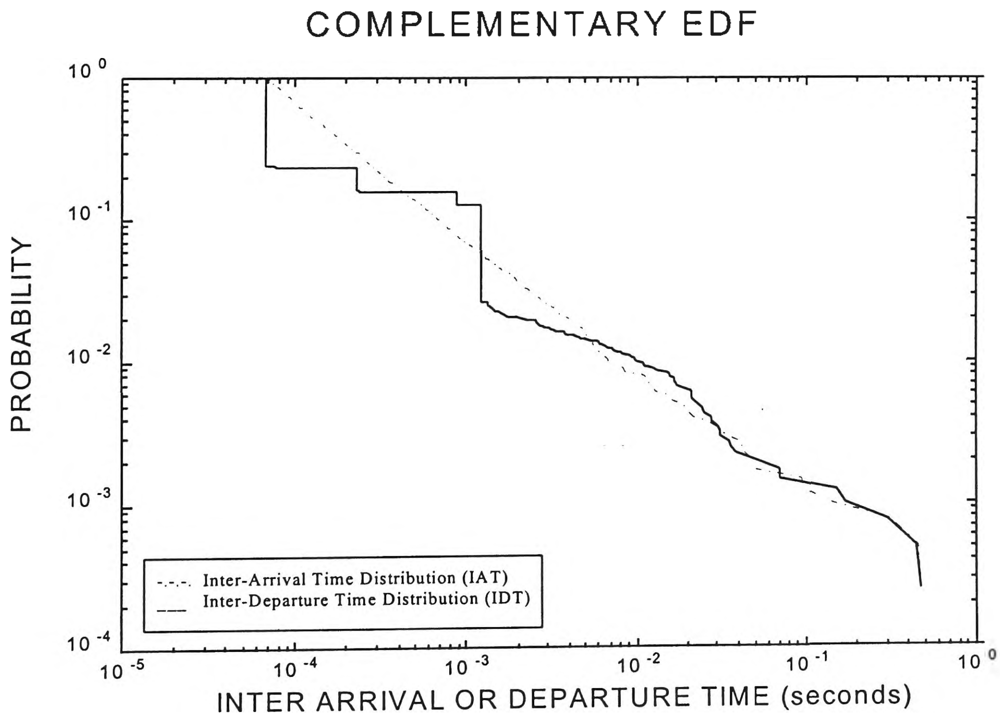


Figure 5.20 Complementary EDF Plot of P30SNLBP1 with $\omega = 0.0000672$

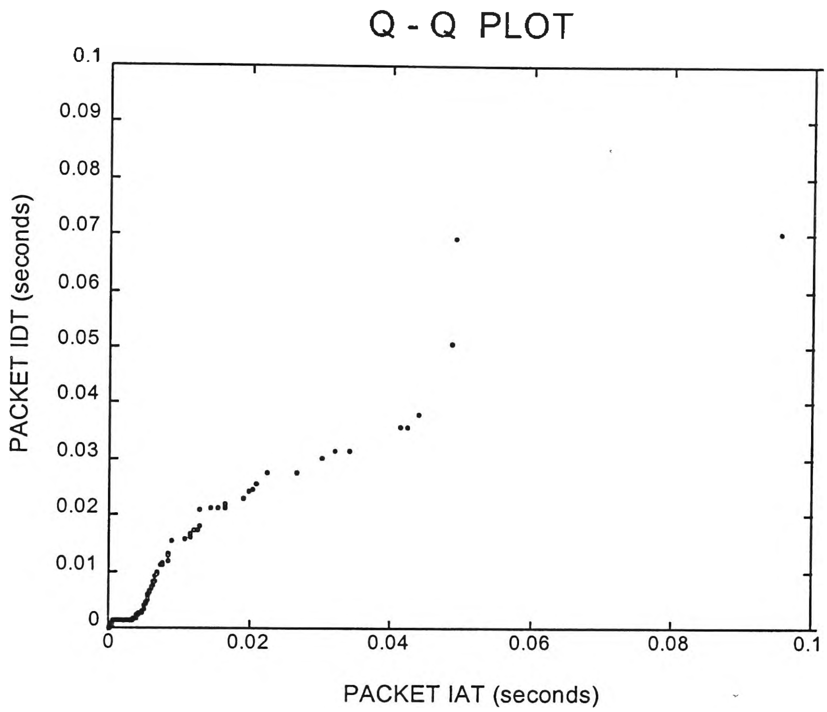


Figure 5.21 Q-Q Plot of P30SNLBP1 with ω = 0.0000672

## 5.7 Summary

This chapter has presented the experimental methodology for testing the null hypothesis and test results. Two traffic source models, the Poisson and the Pareto, are used to test the null hypothesis along with 5 different simulation configurations representing different test parameters. The packet IAT generation technique of each traffic model has also been described. Two packet size distributions, modelling different user parameters, have been selected to represent real packet size distributions as observed in real packet traffic data.

The IDC and VTP techniques used to test the self-similarity in aggregate traffic data have shown that the aggregate traffic data generated by the Pareto traffic sources exhibits self-similar behaviour but is not the case in the Poisson traffic sources. The estimated Hurst parameters of the Pareto aggregate traffic data were found to be related to the shape parameter of the Pareto distribution. As the Pareto shape parameter increases (i.e., as its variability reduces), the estimated Hurst parameters of Pareto aggregate traffic data become small.

In conjunction with the formal methods, two graphical methods, known as EDF plot and Q-Q plot, have been utilised to visually explain the results of formal methods. It has been shown that, compared to the Q-Q plot, the EDF plot is more powerful to reveal any deviation in the tails of the IAT and IDT distributions.

This chapter has achieved the main goals of the thesis, namely testing the validity and investigating the sensitivity of the null hypothesis. The K-S test results conclude that under given conditions of simulation configurations the validity and the sensitivity of the null hypothesis depend on the traffic model used to test the null hypothesis.

In case of Poisson traffic model the K-S test results indicate that the null hypothesis is not sensitive to the network and protocol parameters but sensitive to the user parameters, namely user activity level and packet size distribution. In case of Pareto traffic model the K-S test results show that the null hypothesis is sensitive only to the user parameters, i.e. packet arrival rate and the burstiness of source traffic which is represented by the shape parameter of the Pareto distribution. In both cases the null hypothesis is sensitive to the user activity level.

The null hypothesis is applicable for station with low activity level given that the traffic data is not too bursty. In case of very bursty traffic data, K-S test results indicate that the null hypothesis should be rejected. Since the real traffic data is more likely to be bursty and real packet departure rate is more likely to be unstable it is hard to find the applicability range of the null hypothesis in real conditions.

# 6. LAN Traffic Measurement and Modelling: A Case Study

## 6.1 Introduction

The null hypothesis test results as presented in the previous chapter suggested that the null hypothesis may be applicable provided that the station activity level is low and network utilisation level is also low. The previous chapter also showed that graphical methods used in goodness-of-fit techniques might lead to the wrong conclusions if not used in conjunction with formal numerical techniques. In this chapter we present a case study that is intended to show the liability of applying the null hypothesis in measurement-based traffic modelling study.

In Section 6.2 we review the previous study under consideration. This study applies the null hypothesis based only on the results of graphical methods. In particular, this study proposed a measurement-based traffic model that is able to capture the self-similar processes in real traffic data. We then employee the proposed traffic model in our simulation program to study the results claimed in that study.

To complete our verification of the approach used in the previous study we also collected traffic data from our typical departmental LAN using TCPDUMP. The results are presented in Section 6.3. In this section we first show that our real traffic data also reveals the self-similarity nature as observed in other studies. We then analyse the individual traffic characteristics based on station's packet inter-departure time distribution. We study the packet size distribution as well to complete the traffic description.

Section 6.4 presents the simulation and analysis of simulated traffic data using the proposed traffic model. In particular, we point out two major discrepancies between the simulation results and real traffic characteristics. We also provide explanations of the possible causes of the discrepancies between real traffic data characteristics and simulation traffic data characteristics generated by the proposed measurement-based traffic model. The findings in this chapter are then summarised in Section 6.5.

### 6.2 Previous Study Review

As an object of the case study we review previous measurement-based traffic modelling (Liu, Anido & Chicharo 1994) which has been described earlier in Section 2.3. For the purpose of further discussion we address this study as *the previous study*. Based on the Ethernet traffic data collected from a typical departmental LAN, the previous study proposed a traffic source model which they call Hybrid traffic model (i.e., combination of Poisson and Pareto). This model has also been applied later to ATM traffic modelling for congestion control (Liu, Anido

& Chicharo 1995). Actually this model was originally proposed by Bond (1987) in his theoretical study of burst noise and has also been previously used to model the error performance of a transmission link (Sexton & Reid 1992, pp. 163-169). In their study, Liu, Anido and Chicharo (1994) also claim that their traffic model is able to capture the self-similar behaviour of real traffic data. Unfortunately this claim was not supported by further statistical analysis.

More specifically their proposed traffic model, the Bond distribution, has 2 parameters with the following survival function (i.e., combination of Poisson and Pareto)

$$F(t) = P[T \geq t] = (1-\beta)t^{-\beta}e^{-\lambda t} \qquad \textbf{(Fml. 6.1)}$$

which has finite rate (Bond 1987), i.e.,

$$\textit{Bond Packet Arrival Rate} = \textit{1/mean} = \frac{\lambda^{(1-\beta)}}{\Gamma(2-\beta)} \qquad \textbf{(Fml. 6.2)}$$

where Γ(x) is a gamma function.

As with the Poisson and Pareto traffic model, to generate the Bond packet inter-arrival time (IAT) in our simulation model we employ a transformation technique. Unfortunately, the inverse function of this distribution is quite complex. However, by using the formal mathematical language MAPLE (Heck 1996), we found that the inverse function of the Bond distribution has the form known as the *Omega Function* which is a kind of transcendental function (Fritsch, Shafer & Crowley

1973). The MAPLE solution shows that the IAT of Bond distribution has the following form

$$IAT = \frac{\beta}{\lambda}\omega(X)........where...X = \frac{\lambda}{\beta}(1-\beta)^{1/\beta}U^{-1/\beta} \qquad \textbf{(Fml. 6.3)}$$

where $\omega(X)$ is an omega function and $U$ is uniform continuous variate between 0 and 1. There are two iterative algorithms proposed by Fritsch, Shafer and Crowley (1973) to solve this omega function. For the purpose of our study we use their first algorithm (Version A) since our experiments have shown that the first algorithm is more accurate. Figure 6.1 depicts a typical Bond distribution generated by Formula 6.3 with β = 0.62 and λ = 0.0003.

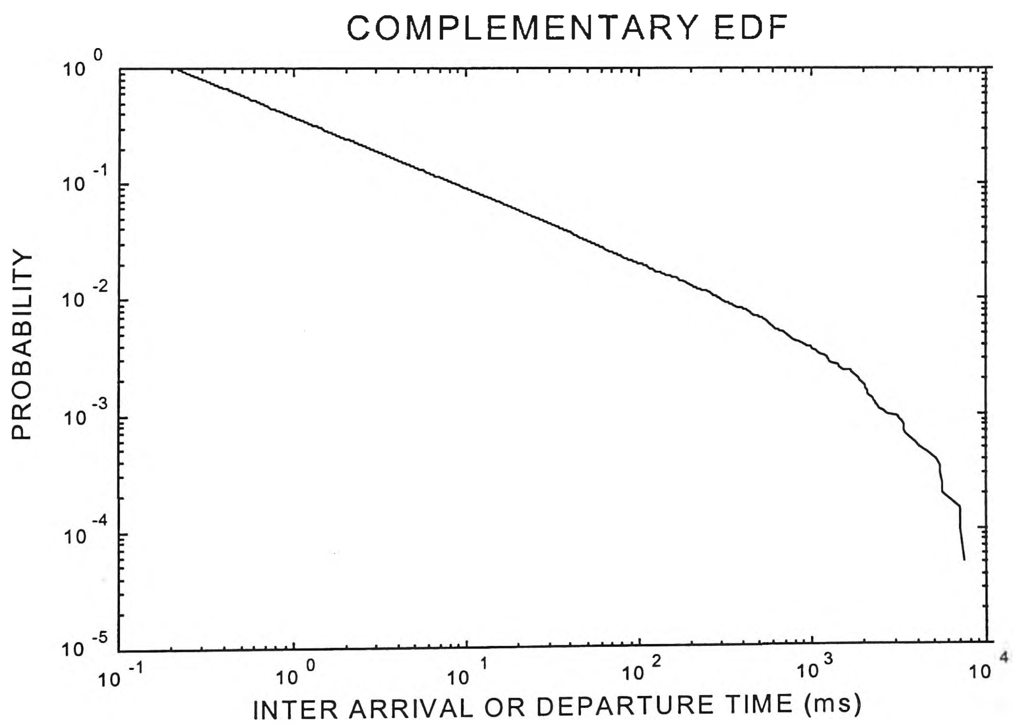


Figure 6.1 Typical Bond Distribution Generated by Formula 6.3

The previous study also indicated that the proposed traffic model parameters vary depending on the average traffic load generated by a station. This metric is comparable to the packet departure rate in our term. They found that the value of parameter $\beta$ increases as the traffic load is increased. In particular they calculate parameter values of their proposed traffic model as shown in Table 6.1.

Table 6.1 Bond Distribution Parameter Values for Different Traffic Loads as Calculated in The Previous Study

| Traffic Load (in Packets/Minute) | Traffic Load (in Packets/Second) | $\beta$ | $\lambda$ |
|---|---|---|---|
| 716 | 11.933 | 0.65 | 0.0003 |
| 230 | 3.833 | 0.62 | 0.0003 |
| 160 | 2.667 | 0.52 | 0.0004 |

### 6.3 Traffic Measurement and Analysis

To complete our case study we collected traffic data from our typical departmental LAN. We used TCPDUMP for the purpose of traffic data collection and basic analysis. We collected 2 data sets of 1,000,000 Ethernet packets each. The packet data contains a time stamp, packet size as well as source-destination address of packets. The statistics of our traffic data is presented in Table 6.2.

As in the previous study, from the aggregate data we extracted individual traffic data and study packet inter-departure time (IDT) distribution of a station by means of histogram technique with 1 ms resolution (bin width). Note that the previous study examined the client-server traffic characteristics instead of considering traffic data characteristics generated by a particular station. However, our preliminary study have shown that the IDT distributions of both data have the same appearance as the Bond distribution. This is quite reasonable since the client-server traffic data can be considered as part of the traffic data generated by a particular station. A typical real packet IDT distribution, exhibiting the appearance of Bond distribution, is presented in Figure 6.2 for complementary EDF (upper tail). The figure does indicate a similar shape of Bond distribution as observed in the previous study.

Table 6.2 Statistics of Measured Traffic Data

| STATISTICS | DATA SET 1 | DATA SET 2 |
|---|---|---|
| Number of Packets | 1,000,000 | 1,000,000 |
| Number of Active Stations | 31 | 23 |
| Network Utilisation | 5.15% | 1.97% |
| Total Packet Load | 195.7 packets/second | 65.6 packets/second |
| Duration of Data Collection | 5010.1 seconds | 5997.8 seconds |

The Data Set 1 and Data Set 2 are those that appear at Table 2.1 as tcpdump3 and tcpdump2 respectively. In this chapter, the Data Set 1 and 2 are mainly used for two purposes. Firstly to show the self-similarity nature in our measured traffic data and secondly to obtain packet size distribution to be used in simulation program. The overlapping packet problem as discussed in Section 2.5, which mostly affects packet IDT statistics, does not affect the analysis results of IDC and VTP techniques used to detect the presence of self-similarity in the aggregate traffic data. It is because the

IDC and VTP methods are based on counting process. See Section 4.4 for detail explanations.

Figure 6.2 also shows Bond-distributed curve of best fit. Since parameter estimation method for Bond distribution is not currently available (Arnold, B.C. 1997, the author of *Pareto Distributions* (1983), pers.comm., 28 January) we then resort to an optimisation technique (Kuester & Mize 1973, pp. 386-398) for the purpose of parameter estimation. This subject will be explored later in Section 6.4.

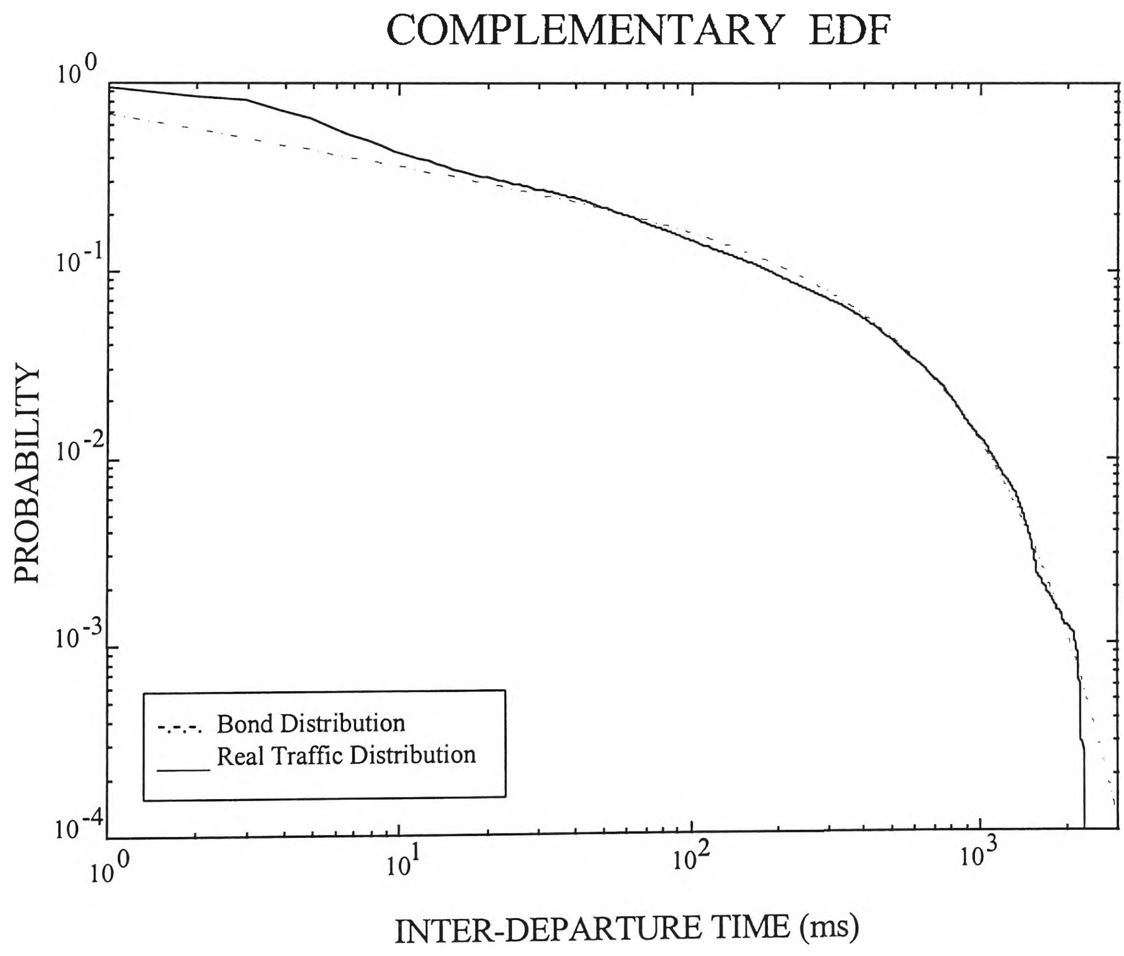


Figure 6.2 Typical Bond Distribution as Observed in Packet's IDT of Real Traffic Data.

### 6.3.1 Self-similarity in Measured Traffic Data

To study the nature of our traffic data we performed IDC and VTP techniques and estimated the Hurst parameter of our data. Figure 6.3 and 6.4 depicts IDC and its related VTP plot of Data Set 1, which has the higher utilisation. The same plots for Data Set 2, having lower utilisation level, are presented in Figure 6.5 and 6.6. The estimated Hurst parameters (0.9307 and 0.9599) indicated that our traffic data is long-range dependent in nature. The last two figures indicated that the self-similar behaviour in packet traffic data is present even at the very low utilisation level.

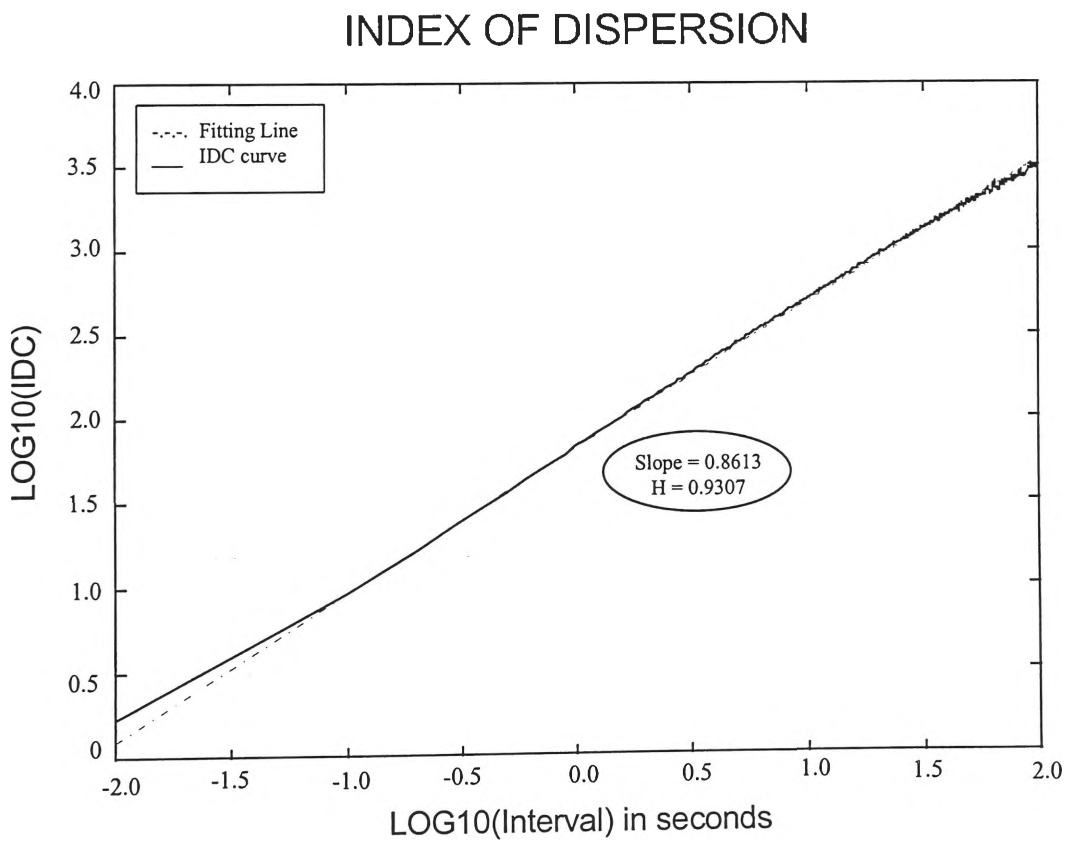


Figure 6.3 IDC Plot of Data Set 1 (Utilisation = 5.15%)

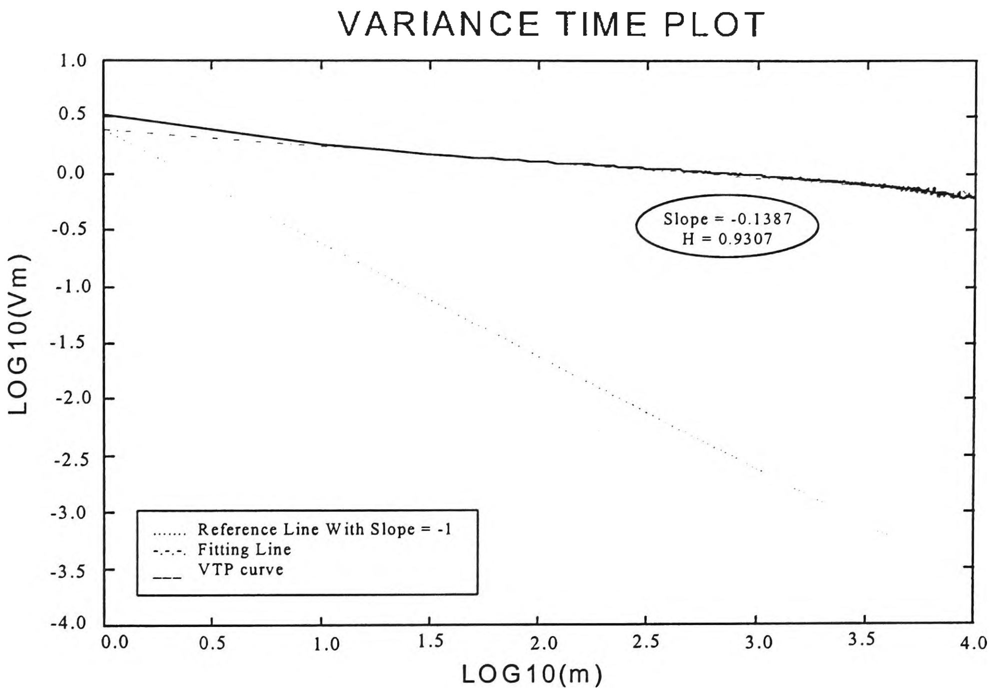


Figure 6.4 VTP Plot of Data Set 1 (Utilisation = 5.15%)

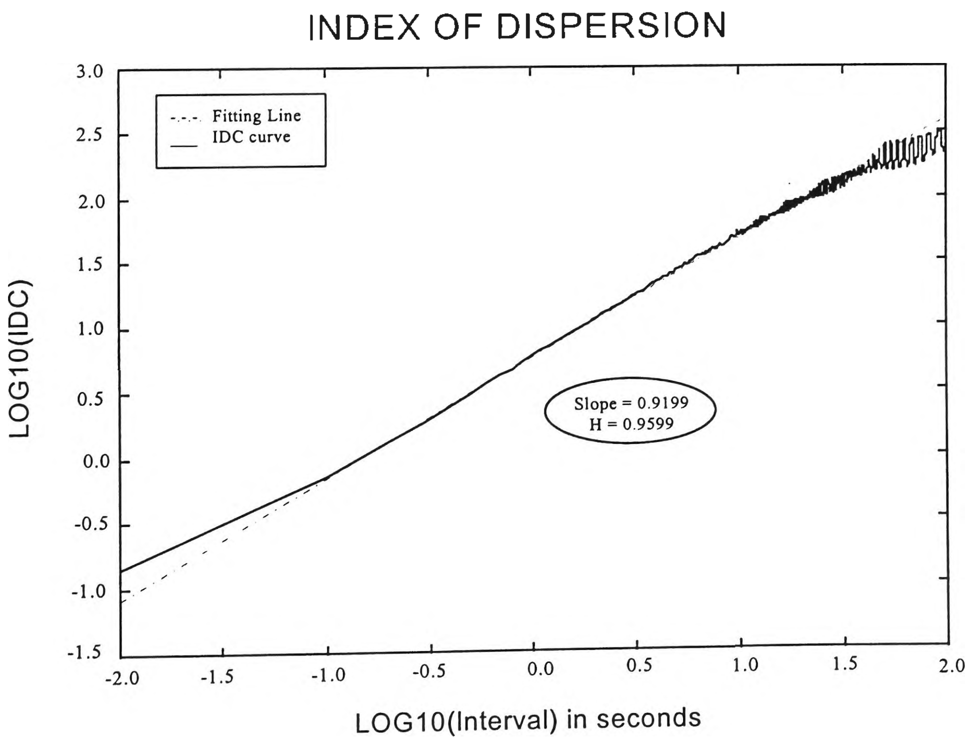


Figure 6.5 IDC Plot of Data Set 2 (Utilisation = 1.97%)

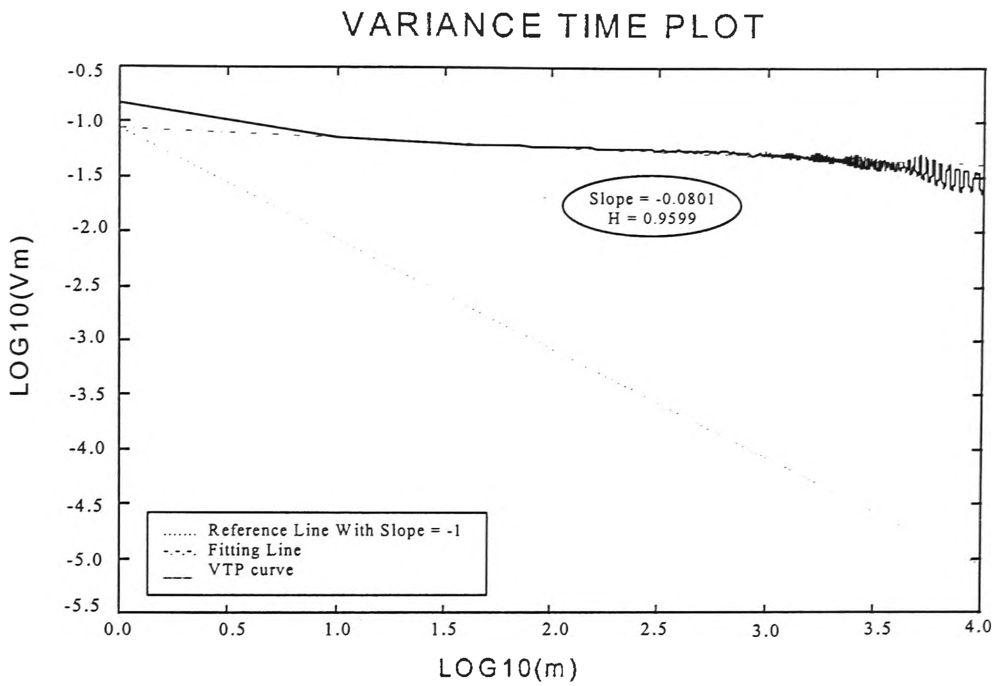


Figure 6.6 VTP Plot of Data Set 2 (Utilisation = 1.97%)

### 6.3.2 Packet Size Distribution

To complete the data description we analysed packet size distribution from our data sets. The results are presented in Figures 6.7 for Data Set 1 and in Figure 6.8 for Data Set 2. The two figures show that packet size distributions of real traffic data consist of many minimum length packets, some maximum length packets and a few of intermediate packet sizes as observed in previous measurement studies (Gusella 1990; Schoch & Hupp 1980; Crane 1981). For the purpose of simulation study in the following section we will use the packet size distribution of P1 as described in Table 5.1 to represent real packet size distribution since its distribution is about the same.

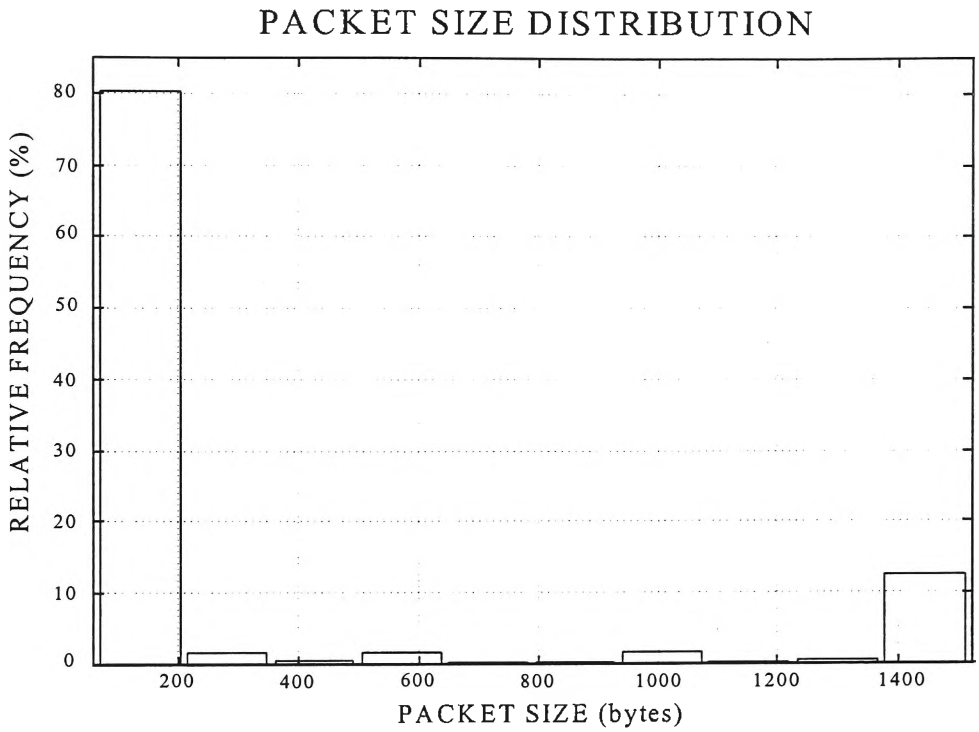


Figure 6.7 Packet Size Distribution of Data Set 1(Utilisation = 5.15%)

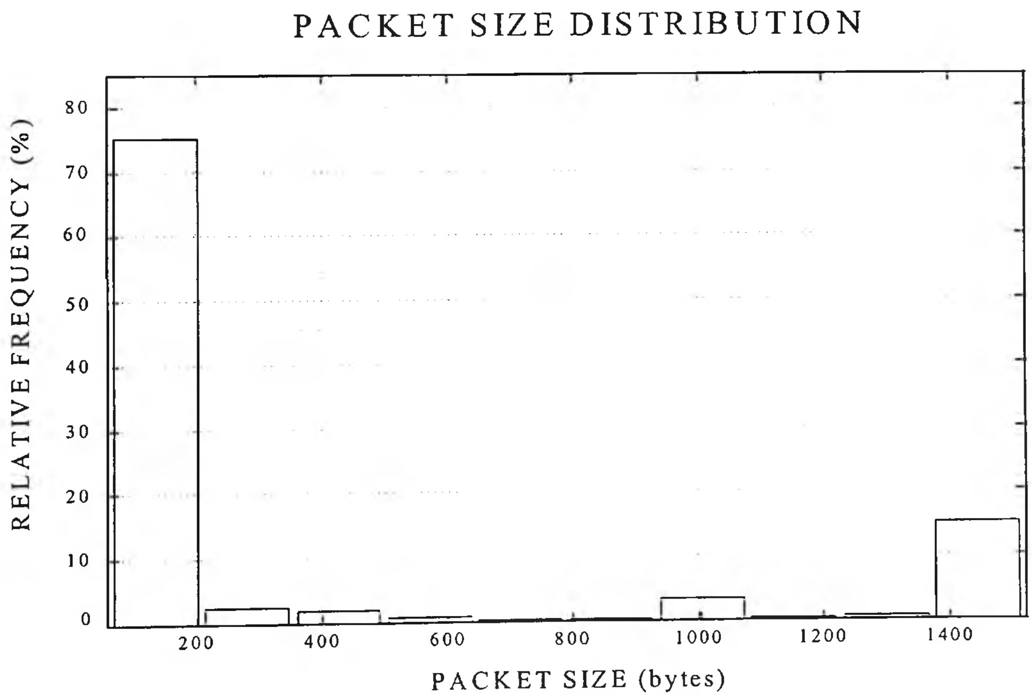


Figure 6.8 Packet Size Distribution of Data Set 2(Utilisation = 1.97%)

## 6.4 Simulation and Analysis

Given the proposed traffic model and information regarding real traffic data characteristics, a simulation study was performed to verify the approach used in the previous study in relation to the null hypothesis. Simulation results were obtained for a reference configuration (with 30 station) and a Bond traffic model.

As in the previous chapter we utilise 3 reference stations to study the null hypothesis. This time we use the parameters proposed in the previous study as in Table 6.1. The parameters of other stations are set to be the same with those in the fourth row of Tables 6.1, namely 0.52 and 0.0004 to guarantee low background traffic. In particular we are interested in 2 statistics as in previous chapter, namely verifying individual station activity levels and self-similar behaviour of aggregate traffic data.

Based on our simulation results, we observed two discrepancies between the measurement results and the simulation results. The first discrepancy observed in our simulation results is in the station activity levels. Our simulation data shows that the packet departure rate of the proposed parameter values (Table 6.1) is significantly different from the observed values of real traffic data. Table 6.3 depicts this discrepancy. These results are consistent within 5 simulation runs with different random seed with total 1,000,000 departed packets for each simulation. The simulated time for each simulation was very long, i.e., about 500,000 seconds. This was due to the very low resultant utilisation level, i.e., 0.04 % with a typical total packet rate of 1.98 packets/second.

Table 6.3 Packet Departure Rate of Simulation and Real Traffic Data for The Proposed Parameter Values

| Real Traffic Load | Simulation Results | Formula 6.2 Calculation |
|---|---|---|
| 11.933 packets/second | 0.0580 packets/second | 0.0656 packets/second |
| 3.833 packets/second | 0.0552 packets/second | 0.0516 packets/second |
| 2.667 packets/second | 0.0278 packets/second | 0.0264 packets/second |

Since the calculation results of Formula 6.2 are close to the simulation results (see Table 6.3), one possible cause for this discrepancy is in the Bond parameter estimation. As mentioned earlier in Section 6.3, the parameter estimation method for Bond distribution is currently not available. To fit real traffic data distribution, as depicted in Figure 6.2, we use the optimisation technique to estimate Bond distribution parameters. The error is mostly in the lower tail of the distribution (see also Figure 6.9). The real traffic data distribution shows that the packet IDT distribution tends to be clustered at low values which indicates the burst nature of packet traffic. The Bond distribution can not capture this behaviour perfectly.

As shown in the previous chapter, the upper tail of IAT and IDT distributions in either two traffic models (Poisson and Pareto) are mostly indistinguishable but is not the case for the lower tail distribution. As an example we refer again to Figure 6.2

which shows the complementary EDF that shows a 'good' agreement between typical IDT data and Bond distribution use to fit that typical data in the upper tail. The data was actually taken from Data Set 1 of a station having an average packet departure rate of 16.8 packets/second. The Bond distribution parameters obtained by optimisation technique are $\beta = 0.31$ and $\lambda = 0.0022$. According to Formula 6.2 the packet rate of this distribution is 0.0162 which is 100 times smaller than expected.

In addition when we apply the K-S test for these distributions, the test result rejects the null hypothesis that the two distributions are the same (the probability is 0.000). Graphical analysis reveals that this rejection is mainly due to the discrepancy in the lower tail. We redraw the EDF of Figure 6.2 as in Figure 6.9 to show this discrepancy.

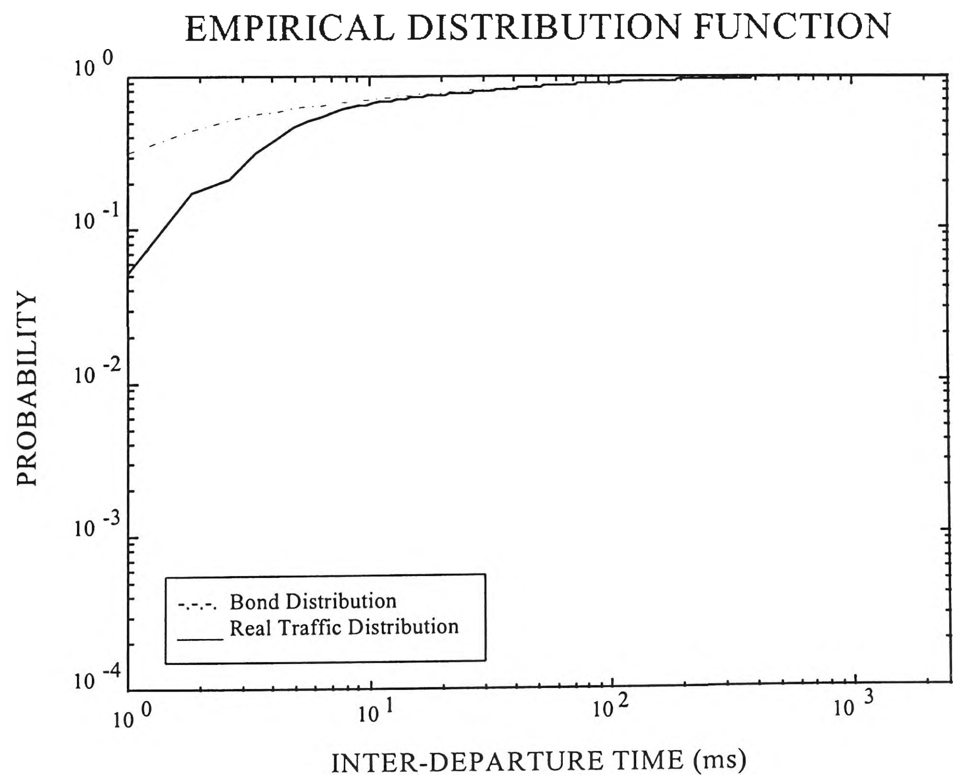


Figure 6.9 EDF Plots Showing Bad Fitting in Lower Tail

Because the real traffic data is bursty in nature, the packet IDT distribution tends to be clustered at lower values, which makes the lower tail distribution an important factor and can not be ignored in traffic modelling.

The second discrepancy is in the self-similar behaviour of aggregate simulation traffic data. The IDC and VTP methods applied to the simulation data have failed to detect the presence of self-similar behaviour in aggregate simulation traffic data. The resulted IDC and VTP values are 0 for this simulation data.

Further investigations were performed using various simulation parameters (user, network and protocol parameters) and various Bond distribution parameter values expecting to obtain different results. However, the characteristics of aggregate simulation data from various configurations are still far from self-similar. A typical IDC and VTP plot is shown in Figure 6.10 and Figure 6.11.

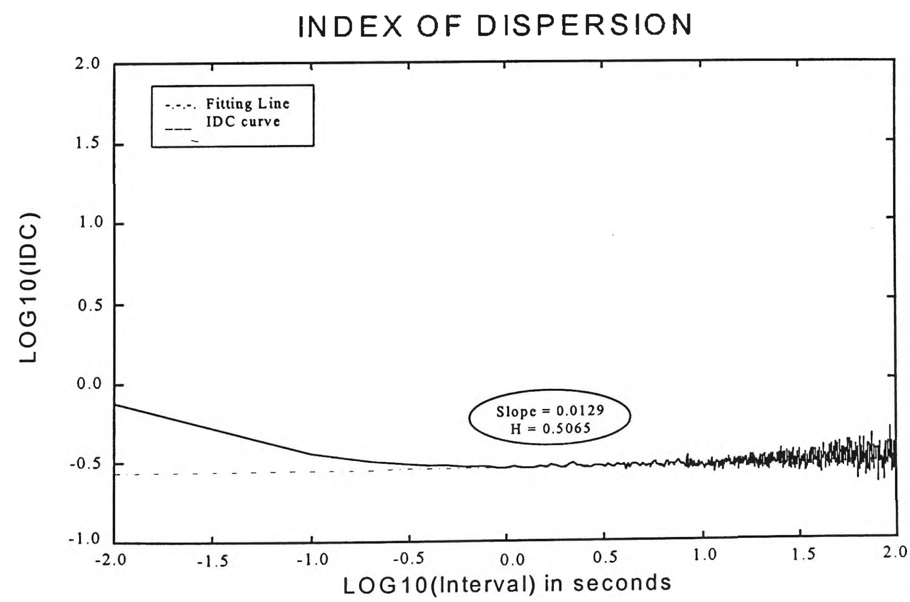


Figure 6.10 Typical IDC Plot of Simulation Data with Bond Traffic Model

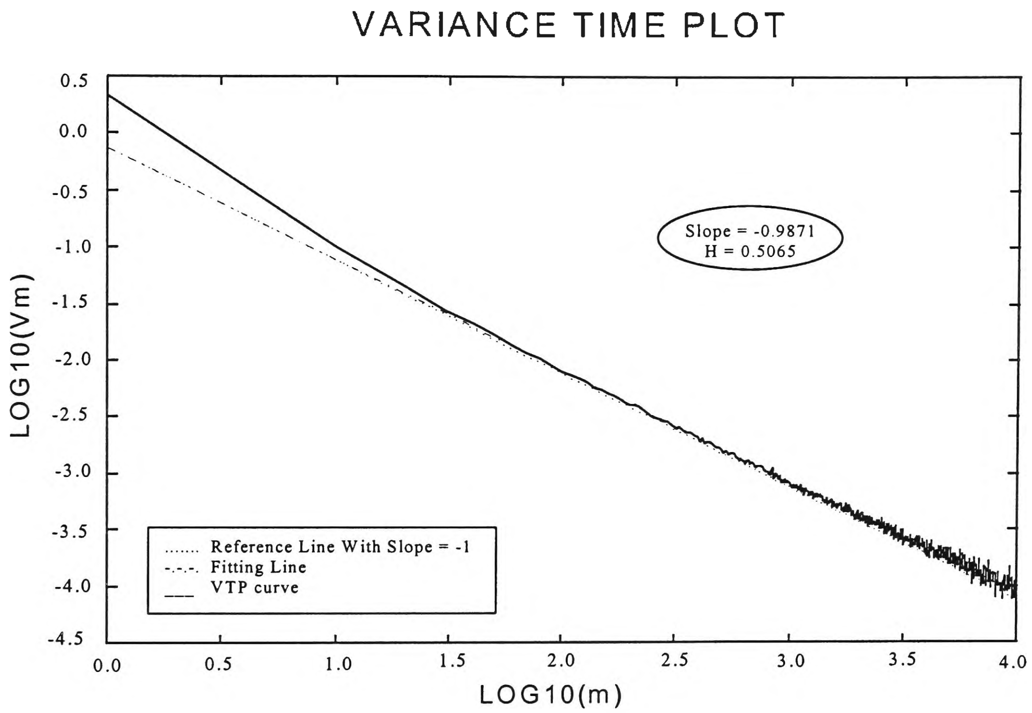


Figure 6.11 Typical VTP Plot of Simulation Data with Bond Traffic Model

The EDF plot of Figure 6.1 and calculation results at Table 6.3 confirm that the simulation program did produce traffic data with Bond distribution properties. On the other hand, the IDC and VTP methods have failed to detect the presence of self-similar behaviour in the aggregate simulation data while the Bond distribution theoretically exhibits self-similar properties. We believe that perhaps there exists certain range of parameter values where the Bond traffic model will generate self-similar traffic data. It is comparable to the Pareto traffic model that will generate self-similar traffic when the shape parameter is less than 2. More research is necessary to fully resolve the issue.

## 6.5 Summary

The case study that shows the liability of applying the null hypothesis in a measurement-based traffic modelling study has been presented. The proposed measurement-based traffic model, known as the Bond distribution, has been used as the object of case study. Two statistics of the resulting simulation traffic data have been considered in particular to show the discrepancies between the characteristics of the real traffic data and the simulation data. The two discrepancies observed in this case study were more likely caused by the problem in Bond parameter estimation. Since the Bond distribution theoretically exhibits self-similar properties, more research is required to fully resolve the issue.

The results of this case study have shown that the null hypothesis approach in traffic modelling should be used with precaution to avoid wrong conclusions. Statistical analysis based only on graphical methods should be avoided since this may lead to the wrong conclusion. Any measurement-based results should be validated either by simulation results or analytical results before they can be justified (Jain 1991, p 32).

# 7. Conclusions

## 7.1 Overview

Traffic modelling plays an important role in telecommunication network design, planning and performance evaluation. To find an accurate and realistic traffic model most traffic modelling studies resort to measurement-based approaches. In this context the measurement results are usually used to validate the traffic model under consideration.

In measurement-based traffic modelling studies some assumptions dealing with network protocol characteristics are usually employed to simplify the study. In case of protocol-specific (individual) traffic modelling the commonly used assumption, as we identified, is that the data link layer protocol (i.e., layer 2 of OSI Reference Model) does not affect the traffic characteristics from its higher layer significantly. In other words, this hypothesis assumes that the traffic characteristics at the input of data link layer protocol is statistically the same as the traffic characteristics at its output, i.e., the physical layer traffic where the measurement can be easily done. This assumption is then addressed as the null hypothesis in this thesis.

There are several factors that may affect the validity of the null hypothesis as argued in Section 2.4, i.e., network parameters, protocol parameters and user parameters. In addition, the accuracy of the measurement systems used to collect traffic data may contribute some serious errors that affect the validity of the derived traffic models based on the null hypothesis as demonstrated in Section 2.5. These matters have led to the interest of testing the validity of the null hypothesis and to find out its sensitivity against several parameters. These subjects are the main areas of concern of this thesis.

Thus the work in this thesis has been mainly concerned with testing the validity and sensitivity of the null hypothesis. The simulation technique is used as a tool to achieved these purposes. Several statistical methods were used to test the null hypothesis. In addition, a case study has been presented to demonstrate the liability of applying the null hypothesis in measurement-based study without further formal validation. Our results and recommendations are described in the following sections, as well as further areas of research.

## 7.2 Validity and Sensitivity of The Null Hypothesis

Two source traffic models, i.e., Poisson and Pareto, were used to study the validity and sensitivity of the null hypothesis. The Poisson traffic model resembled a pure traffic assumption as commonly used in the literature and analytical studies while the Pareto traffic model represented a more realistic traffic model since its aggregate traffic stream can capture self-similar behaviour or long-range dependent process as observed in real traffic data.

For the Poisson traffic model, the K-S test results showed that the null hypothesis is valid for relatively low and medium level activity stations. For low activity stations, the null hypothesis is valid up to 75% utilisation level while for medium activity stations the null hypothesis is valid up to 30% utilisation level. The null hypothesis is not valid for relatively high activity stations.

The K-S test results have also revealed that, for the Poisson traffic model, the null hypothesis is sensitive to the user parameters under consideration, namely station activity level and packet size distributions but not sensitive to network parameters or protocol parameters under consideration.

For the Pareto traffic model the K-S test results showed that the null hypothesis is more likely to be rejected compared to Poisson traffic model results. Compared to Poisson traffic model results, the null hypothesis with Pareto traffic model is sensitive only to the user parameter of activity level but not to the packet size distribution. More specifically, the null hypothesis is sensitive to the burstiness of station as expressed by the shape parameters of Pareto distribution. As the burstiness or the variability of packet arrival processes increases, indicated by lower value of Pareto shape parameter, the null hypothesis is more likely to be rejected.

The IDC and VTP techniques used to test the characteristics of aggregate traffic of simulation data have shown that the aggregate simulation traffic data of Pareto traffic model is more closer to real traffic characteristics as compared to simulation traffic data of Poisson traffic model. The aggregate traffic data of Pareto traffic model was shown to exhibit self-similar behaviour as observed in real traffic data characteristics. Therefore, by comparing the null hypothesis test results of Poisson and Pareto traffic model we conclude empirically that the null hypothesis is valid for relatively low-level activity station (up to 10 packets/second) and sensitive only to the user parameter of user activity level.

However, since the real traffic data could be more bursty in nature and real station's packet departure rate is more likely to be unstable (Smith & Kain 1991), this makes it hard to generalise our results to real conditions.

On the basis of the results demonstrated in Chapter 6, a case study, it is concluded that the application of the null hypothesis based on graphical analysis of the packet inter-departure time distribution without further formal validation is unjustified. To be justified, any measurement-based study results should be validated by analytical or simulation studies.

## 7.3 Future Work

To understand the real effect of data link protocol on the null hypothesis, the study in this thesis could be extended to include more extensive measurement-based study using a dedicated station monitoring system. Such a station will need to implement special code within its data link protocol code to include a probing capability. Since the presence of this code may affect the real traffic characteristics at the input of data link layer protocol, the code should be carefully designed to minimise this effect. By comparing the traffic characteristics measured by the probing code and the traffic characteristics at the output of data link protocol, with and without the presence of probing code, we then can understand the real effect of data link protocol on the real traffic characteristics from its higher layer. The traffic characteristics generated by the monitoring station can be controlled to study the effect of other parameters on the null hypothesis.

In case of traffic source modelling, the approach described above can be used to obtain more accurate traffic characteristics of individual network user independent of data link protocol being used. Since future high speed networks are more likely to use another data link protocol, the resulting accurate traffic source model is very useful to study the performance of future high speed network (e.g., ATM LAN). The study of the resulting traffic source model characteristics in relation to the self-similar behaviour of aggregate traffic data is also another interesting area of future research.

# References

Addie, R.G., Zukerman, M., & Neame, T. 1995, 'Fractal Traffic : Measurements, Modelling and Performance Evaluation', in *Proceedings IEEE INFOCOM'95.*

Al-Salqan, Y.Y., Kang, C., Herzog, J.H. & Park, E.K. 1991, 'A Simulation Study of a CSMA/CD with Connected Data Links', in *Proceedings of the 24th Annual Simulation Symposium*, pp. 292-300, IEEE Compt. Soc. Press, Los Alamitos.

Archie, K., Campbell, G., Cathey, G., Cohen, A. & Finn, R. 1993, *OPNET Modeller Manual 8.1.0, Example Models Manual/Protocol Models vol.0*, pp. Eth-3-Eth.-16, MIL 3, Inc., Washington DC.

Arlitt, M.F. & Williamson, C.L. 1996, 'Web Server Workload Characterisation: The Search for Invariants (Extended Version)', pre-print.

Arnold, B.C. 1983, *Pareto Distributions*, International Co-operative Publishing House, Fairland, MD.

Basu, S., Mukherjee, A. & Klivansky, S. 1996, 'Time Series Models for Internet Traffic', in *Proceedings of INFOCOM'96*, vol. 2, pp. 611-620.

Black, Uyless 1993, *Computer Networks : Protocol, Standard, Interface*, 2nd edn, Prentice Hall, Englewood Cliffs.

Boggs, D.R., Mogul, J.C. & Kent, C.A. 1995, 'Measured Capacity of an Ethernet: Myths and Reality', *Computer Communication Review*, vol. 25, no. 1, pp. 124-137.

Bond, D.J. 1987, 'A Theoretical Study of Burst Noise', *British Telecommunications Technology Journal*, vol. 5 no. 4, pp. 51-60.

Cáceres, R., Danzig, P.B., Jamin, S. & Mitzel, D.J. 1991, 'Characteristics of Wide-Area TCP/IP Conversations', *Proc. of ACM SIGCOMM'91*, pp. 101-112.

Chao, H.J., Ghosal, D., Saha, D. & Tripathi, S.K. 1994, 'IP on ATM Local Area Networks', *IEEE Communications Magazine*, August, pp. 52-59.

Cinotti, M., Giordano, S., Romani, F. & Russo, F. 1995, 'Implications of the Self-Similar Behaviour of Real Traffic Sources on the Local-Queue Occupancy in Metropolitan Area Networks', in *Proc. of IEEE Singapore International Conference on Networks*, Singapore, July. Available : UMI File: IPO (IEEE/IEE Publications Ondisc).

Cinotti, M., Mese, E.D., Giordano, S. & Russo, F. 1994, 'Long-Range Dependence in Ethernet Traffic Offered to Interconnected DQDB MANS', Paper presented to Singapore ICCS'94, Singapore, November. Available : UMI File: IPO (IEEE/IEE Publications Ondisc).

Comer, D.E. 1991, *Internetworking with TCP/IP*, volume 1, 2nd edn, Prentice Hall, Englewood Cliffs.

Crane R.C. 1981, 'Software Pack and Controller Link DEC Computers in an Ethernet', *Electronics*, December.

D'Agostino, R.B & Stephens, M.A (eds) 1986, *Goodness-Of-Fit Techniques*, Marcel Dekker Inc., New York.

Danzig, P.B., Jamin, S., Caceres, R., Mitzel, D.J. & Estrin, D. 1992, 'An Empirical Workload Model for Driving Wide-Area TCP/IP Network Simulations', *Internetworking : Research and Experience*, vol. 3, no. 1, pp. 1-26.

Deng, S, Bugos, A.R. & Hill, P.M. 1996, 'Design and Evaluation of an Ethernet-Based Residential Network', *IEEE JSAC*, vol. 14, no. 6, pp. 1138-1150.

Devai, J.J., Kerrigan, T.C. & Molloy, M.K. 1990, 'Simulation of The LAN Behaviour in The Distributed HP-UX Environment', in *Proceedings of the 1990 Winter Simulation Conference*, pp. 805-812, IEEE Compt. Soc. Press, Los Alamitos.

Duffield, N.G., Lewis, J.T., O'Connell, N., Russell, R. & Toomey, F. 1994, 'Statistical issues raised by the Bellcore data', in *Proceedings of the IEE Eleventh UK Teletraffic Symposium*, pp. 1-8, IEE, London.

Erramilli, A., Narayan, O. & Willinger, W. 1996, 'Experimental Queuing Analysis with Long-Range Dependent Packet Traffic', *IEEE/ACM Transaction on Networking*, vol. 4, no. 2, pp. 209-223.

Evans, M., Hastings, N. & Peacock, B. 1993, *Statistical Distributions*, 2nd edn., John Wiley & Sons, New York.

Fisher, N.I. 1983, 'Graphical Methods in Nonparametric Statistics: A Review and Annotated Bibliography', *International Statistical Review*, vol. 51, pp. 25-58.

Fishman, G.S. 1978, *Principles of Discrete Event Simulation*, John Wiley & Sons, New York.

Fowler, H.J., Leland, W.E. 1991, 'Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management', *IEEE JSAC*, vol. 9, no. 7, pp. 1139-1149.

Fritsch, F.N., Shafer, R.E. & Crowley, W.P. 1973, 'Solution of the Transcendental Equation $we^{w} = x$', *Communication of the ACM*, vol. 16, no. 2, pp. 123-124.

Frost, V.S. & Melamed, B. 1994, 'Traffic Modelling For Telecommunications Networks', *IEEE Communications Magazine*, March, pp. 70-81.

Georganas, N.D. 1994, 'Self-Similar ("Fractal") Traffic in ATM Networks', in *Proceedings of The 2nd International Workshop on Advanced Teleservices and High Speed Communication Architectures (IWACA'94),* pp. 1-7.

Gerson, M. 1975, 'The Techniques and Uses of Probability Plotting', *The Statistician*, vol. 24, no. 4, pp. 235-257.

Gnanadesikan, R. 1977, *Methods for Statistical Data Analysis of Multivariate Observations*, John Wiley & Sons, New York.

Gonsalves, T.A. 1987, 'Measured Performance of the Ethernet', in *Advances in Local Area Networks,* eds K. Kümmerle, F.A. Tobagi & J.O. Limb, IEEE Press, New York.

Gonsalves, T.A. & Tobagi, F.A. 1988, 'On the Performance Effects of Station Locations and Access Protocol Parameters in Ethernet Networks', *IEEE Transactions On Communications*, vol. 36, no. 4, pp. 441-449.

Gordon, J.J. 1995. 'Long Range Correlation in Multiplexed Pareto Traffic', in *Proceedings of The International IFIP-IEEE Conference on Broadband Communications 1996*, Chapman & Hall, London, pp. 28-39.

Gusella, Riccardo. 1990, 'A measurement study of diskless workstation traffic on an Ethernet', *IEEE Transactions On Communications*, vol. 38, no. 9, pp. 1557-1568.

Gusella, Riccardo. 1991, 'Characterising the Variability of Arrival Processes with Index of Dispersion', *IEEE JSAC*, vol. 9, no. 2, pp. 203-211.

Hampel, F. 1987, 'Data Analysis and Self-Similar Processes', in *Bulletin of The International Statistical Institute, Proceedings of the 46th Session*, Book 4, vol. LII, pp. 235-254.

Hayes, M. & Molle, M.L. 1995, 'Solving Capture in Switched Two-Node Ethernet by Changing Only One Node', in *Proceedings of the 20th Conference on Local Computer Networks*, pp. 387-395, IEEE Compt. Soc. Press, Los Alamitos.

Heck, A. 1996, *Introduction to MAPLE*, 2nd edn, Springer Verlag, New York.

Hsu, J.Y. 1996, *Computer Networks: Architecture, Protocols, and Software*, Artech House, Boston.

Huang, J.H. & Chen, C.W. 1992, 'On Performance Measurements of TCP/IP and Its Device Driver' in *Proceedings of the 17th Conference on Local Computer Networks*, pp. 568-575, IEEE Compt. Soc. Press, Los Alamitos.

Hunt, Craig 1992, *TCP/IP Network Administration*, O'Reilly & Associates, Inc., Sebastopol, California.

Jacobson, V., Leres, C. & McCanne, S. 1994, TCPDUMP manual, available : UNIX on line manual.

Jain, Raj 1991, *The Art of Computer Systems Performance Analysis, Techniques for Experimental Design, Measurement, Simulation and Modeling*, John Wiley & Sons Inc., New York.

Jain, Raj. & Routhier, Shawn A. 1986, 'Packet Trains - Measurements and a New Model for Computer Network Traffic', *IEEE JSAC*, vol. 4, no. 6, pp. 986-995.

Johnson, D.H. & Kumar, A.R. 1990. 'Modeling and Analyzing Fractal Point Processes', in *Proceedings of IEEE Int. Conf. on Acoust., Speech and Signal Proc. ICASSP-90*, pp. 1353-1356.

Johnson, N.L., Kotz, S. & Balakrishnan, N. 1994, *Continuous Univariate Distributions*, Volume 1, 2nd edn, John Wiley & Sons, New York.

Kao, Shang-Juh 1993, 'Variations on Ethernet Traffic Monitoring', in *Proceedings IEEE TENCON'93*, IEEE Compt. Soc. Press, Los Alamitos.

Khalil, K.M., Luc, K.Q. & Wilson, D.V. 1990, 'LAN Traffic Analysis and Workload Characterization' in *Proceedings of the 15th Conference on Local Computer Networks*, pp. 112-122, IEEE Compt. Soc. Press, Los Alamitos.

Kotz, S. & Johnson, N.L. 1985, 'Probability Plotting', in *Encyclopaedia of Statistical Sciences*, John Wiley & Sons, New York, vol. 7, pp. 232-237.

Kuester, J.L. & Mize, J.H. 1973, *Optimization Techniques with FORTRAN*, McGraw Hill, New York.

Lam, S.S. 1980, 'A carrier sense multiple access protocol for local networks', *Computer Networks*, vol. 4, February, pp. 21-32.

Law, A.M. & Kelton, W.D. 1991, *Simulation Modeling and Analysis*, 2nd edn, McGraw-Hill Inc., New York.

Leffler, S.J., McKusick, M.K., Karels, M.J. & Quarterman, J.S. 1989, *The Design and Implementation of the 4.3BSD UNIX® Operating System*, Addison-Wesley, Reading.

Leland, W.E. & Wilson, D.V. 1991, 'High Time-Resolution Measurement and Analysis of LAN Traffic : Implications for LAN Interconnection', *IEEE INFOCOM'91*, pp. 1360-1366.

Leland, W.E., Taqqu, M.S., Willinger, W. & Wilson, D.V. 1993, 'On the Self-Similar Nature of Ethernet Traffic' , in *Proc. of The ACM SIGCOMM'93*, pp. 183-193.

_____ 1994, 'On the Self-Similar Nature of Ethernet Traffic (extended version)', *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1-15.

Leon-Garcia, Alberto 1994, *Probability and Random Processes for Electrical Engineering*, 2nd edn, Addison-Wesley, Reading.

Liu, T., Anido, G. & Chicharo, J. 1994, 'Local Area Network Traffic Characterisation and Modelling', in *Proceedings of Australian Telecommunication Networks & Applications Conference*, vol. 1, pp. 335-340.

_____ 1995, 'ATM Traffic Modelling for Congestion Control', in *Proceedings of Australian Telecommunication Networks & Applications Conference*, vol. 1, pp. 91-96.

Matsumoto, Y., Takashi, Y. & Hasegawa, T. 1990a, 'Probability Distributions of Interdeparture and Response Time in Multipacket CSMA/CD Systems', *IEEE Transactions On Communications*, vol. 38, no. 1, pp. 54-66.

_____ 1990b, 'The Effects of Packet Size Distributions on Output and Delay Processes of CSMA/CD', *IEEE Transactions On Communications*, vol. 38, no. 2, pp. 199-214.

Mazraani, T.Y. & Parulkar, G.M. 1992, 'Performance Analysis of The Ethernet Under Conditions of Bursty Traffic', in *Proceedings of the IEEE GLOBECOM'92.*

McCord, J.W. 1992, *Borland C++ 3.1 Programmer's Reference*, 2nd edn. Que Corporation, Carmel.

Metcalfe, R.M. & Boggs, D.R. 1976, 'Ethernet: Distributed packet switching for local computer networks', *Communication of The ACM*, vol. 19, pp. 395-404.

Minnich, N.M. & Cotton, C.J. 1983, 'An Evaluation of Two Unibus Ethernet Controller', in *Proceedings of The 8th Conference on Local Computer Networks*, IEEE Computer Society, Los Angeles, pp. 29-36.

Morino, Y. & Takahara, M. 1992, 'Measurement of Packet Generation and Communication Characteristics in Ethernet System', in *Proceedings of the Singapore ICCS/SITA'92*, IEEE Compt. Soc. Press, Los Alamitos.

Newman, Peter 1994b, 'Traffic Management for ATM Local Area Networks', *IEEE Communications Magazine*, August, pp. 44-50.

OPNET 1993, *OPNET Modeling Manual*, Release 2.4, MIL 3 Inc., Washington DC.

Papadopoulos, C. 1992, 'Remote visualisation on campus network', M.S. Thesis, Dept. Comp. Sci., Washington University, St. Louis, MO.

Papadopoulos, C. & Parulkar, G.M. 1993, 'Experimental Evaluation of SunOS IPC and TCP/IP Protocol Implementation', *IEEE/ACM Transactions On Networking*, vol. 1, no. 2, pp. 199-216.

Pawlita, P.F. 1989, 'Two Decades of Data Traffic Measurements: A Survey of Published Results, Experiences and Applicability', in *Teletraffic Science for New Cost-Effective Systems, Networks and Services, ITC-12*, ed. M. Bonatti, Elseiver, North-Holland, pp. 230-238.

Paxson, V., Floyd, S. 1994, 'Wide-Area Traffic : The Failure of Poisson Modeling', in *Proc. of The ACM/SIGCOMM'94*, pp. 257-268.

Paxson, V. 1995, 'Fast Approximation of Self-Similar Network Traffic', pre-print.

Porter III, J.E., Coleman, J.W. & Moore, A.H. 1992, 'Modified KS, AD, and C-vM Tests for the Pareto Distribution with Unknown Location & Scale Parameters', *IEEE Transaction on Reliability*, vol. 41, no. 1, pp. 112-117.

Prasad, K. & Patel, R. 1988, 'Performance Analysis of Ethernet Based on an Event Driven Simulation Algorithm', in *Proceedings of the 13th Conference on Local Computer Networks*, pp. 253-267, IEEE Compt. Soc. Press, Los Alamitos.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. 1994, *Numerical Recipes in C, The art of Scientific Computing*, 2nd edn., Cambridge University Press, Cambridge.

Ramakrishnan, K.K. & Yang, Henry 1994, 'Ethernet capture effect: Analysis and Solution', in *Proceedings of the 19th Conference on Local Computer Networks*, pp. 228-240, IEEE Compt. Soc. Press, Los Alamitos.

Ripley, B.D. 1983, 'Computer Generation of Random Variables: A Tutorial', *International Statistical Review*, vol. 51, pp. 301-319.

Rives, T.B., Settle, G., Jobe, S. & Ingles, F. 1988, 'Simulation of The NASA HOSC Ethernet Peripheral Processing Network', in *Proceedings of The 1988 IEEE South East Conference*, IEEE Soc. Press, New York.

Robert, S. & Lambrecht, C.B. 1995, 'An SSMP Traffic Model Based on Network Measurements', in *Proc. of The 1995 IEEE 14th Annual International Phoenix Conference on Computers and Communications*, pp. 494-500, IEEE Compt. Soc. Press, Los Alamitos.

Robert, S. & Boudec, J.Y.L. 1996, 'Can self-similar traffic be modeled by Markovian processes ?', in *Lecture Notes in Computer Science, volume 1044 (Proc. 1996 Int. Zurich Sem. on Dig. Comm.)*, ed. Bernhard Plattner, Springer, pp. 119-130.

Ryu, B.K. 1996, 'Implications of Self-Similarity for Providing End-to-End QOS Guarantees in High-Speed Networks : A Framework of Application Level Traffic Modeling', in *Lecture Notes in Computer Science, volume 1044 (Proc. 1996 Int. Zurich Sem. on Dig. Comm.)*, ed. Bernhard Plattner, Springer, pp. 65-79.

Ryu, B.K. & Lowen, S.B. 1996, 'Point Process Models for Self-Similar Network Traffic, with Applications', pre-print 1996, to appear in Stochastic Models 1997.

Sadiku, M.N.O. & Ilyas, M. 1995, *Simulation of Local Area Networks*, CRC Press, Boca Raton, Florida.

Schwartz, Mischa 1987, *Telecommunication Networks: Protocols, Modeling and Analysis*, Addison-Wesley, Reading, Massachusetts.

Senior, J.M., Rehal, D.S. & Wiseman, A.N. 1992, 'Efficient Traffic Model to Obtain LAN Performance Under Asymmetric Traffic Loads', in *Proceedings of The Singapore ICCS/ISITA '92*, IEEE.

Sexton, M. & Reid, A. *1992, Transmission Networking : SONET and The SDH*, Artech House, London.

Shoch, J.F., Dalal, Y.K., Redell, D.D. 1982, 'Evolution of The Ethernet Local Computer Network', *Computer*, August, pp. 10-27, IEEE.

Shoch, J.F. & Hupp, J.A. 1980, 'Measured Performance of an Ethernet Local Network', *Communications of the ACM*, vol. 23, no. 12, pp. 711-721.

Smith, W.R. & Kain, R.Y. 1991, 'Ethernet Performance under Actual and Simulated Loads', in *Proceedings of the 16th Conference on Local Computer Networks*, pp. 569-581, IEEE Compt. Soc. Press, Los Alamitos.

SNOOP manual, available : UNIX on line manual.

Sohraby, K., Molle, M.L. & Venetsanopoulos, A.N. 1987, 'Comments on "Throughput Analysis for Persistent CSMA Systems"', *IEEE Transactions On Communications*, vol. COM-35, no. 2, pp. 240-243.

Stallings, William 1994, *Data and Computer Communications*, 4th edn, Macmillan, Englewood Cliffs.

Stamoulis, G.D., Anagnostou, M.E. & Georgantas, A.D. 1994, 'Traffic source models for ATM networks: a survey', *Computer Communications*, vol. 17, no. 6, pp. 428-438.

Stevens, W.R. & Wright, G.R. 1995, *TCP/IP Illustrated: The Implementations*, vol. 2, Addison-Wesley, Reading, Massachusetts.

Takagi, H. & Kleinrock, L. 1987, 'Correction to "Throughput Analysis for Persistent CSMA Systems"', *IEEE Transactions On Communications*, vol. COM-35, no. 2, pp. 243-245.

Takashi, Y., Matsumoto, Y. & Hasegawa, T. 1986, "Probability Distributions of Delay and Interdeparture Time in Non-slotted CSMA/CD", in *Local Communication Systems : LAN and PBX*, J.P. Eds. Cabamel et al., North-Holland, pp. 423-435.

Tan, H.H. & Tsai, K. 1996, 'Packet Output Processes of CSMA and CSMA/CD Protocols', *IEEE Transactions On Communications*, vol. 44, no. 4, pp. 464-474.

Tanenbaum, A.S. 1988, *Computer Networks*, 2nd edn, Prentice-Hall, Englewood Cliffs.

Taqqu, M.S., Teverovsky, V. & Willinger, W. 1995, 'Estimators for long-range dependence: an empirical study', *Fractals*, vol. 3, no. 4, pp. 785-788.

Taylor, D., Oster, D.L. & Green, L. 1983, 'VLSI Node Processor Architecture for Ethernet', *IEEE JSAC*, vol. SAC-1, no. 5, pp. 733--739.

Tobagi, F.A. 1982, "Distributions of Packet Delay and Interdeparture Time in Slotted ALOHA and Carrier Sense Multiple Access", *Journal of The ACM*, vol. 29, no. 4, pp. 907-927.

Tobagi, F.A. & Hunt, V.B. 1987, 'Performance Analysis of Carrier Sense Multiple Access with Collision Detection', in *Advances in Local Area Networks,* eds K. Kümmerle, F.A. Tobagi & J.O. Limb, IEEE Press, New York.

Wilk, M.B. & Gnanadesikan, R. 1968, 'Probability Plotting Methods for The Analysis of Data', *Biometrika*, vol. 55, no. 1, pp. 1-17.

Willinger, W., Taqqu, M.S. & Erramilli, A. 1996, 'A bibliographical guide to self-similar traffic and performance modeling for modern high-speed networks', in *Stochastic Networks, Theory and Applications*, Eds F.P. Kelly, S. Zachary and I. Ziedins, Oxford University Press, New York.

Willinger, W., Taqqu, M.S., Leland, W.E. & Wilson, D.V. 1995, 'Self-Similarity in High-Speed Packet Traffic : Analysis and Modeling of Ethernet Traffic Measurements' *Statistical Science*, vol. 10, no. 1, pp. 67-85.

Willinger, W., Taqqu, M.S., Sherman, R. & Wilson, D. 1995, 'Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level (Extended Version), pre-print.

Wolfberg, N.E. (ed.) 1983, *The Ethernet Handbook*, Shotwell and Associates, Portola Valley, California.